Analiza svojstava raspada snopova korištenjem neuralne mreže u ulozi klasifikatora

Jerčić, Ivan

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

Ivan Jerčić

Analiza svojstava raspada snopova korištenjem neuralne mreže u ulozi klasifikatora

Diplomski rad

Zagreb, 2021.

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI STUDIJ
FIZIKA; SMJER ISTRAŽIVAČKI

Ivan Jerčić

Diplomski rad

Analiza svojstava raspada snopova korištenjem neuralne mreže u ulozi klasifikatora

Voditelj diplomskog rada: izv. prof. dr. sc. Nikola Poljak

Ocjena diplomskog rada: _____

Povjerenstvo: 1. _____

2. _____

3. _____

Datum polaganja: 19.07.2021.
Zagreb, 2021.

Sažetak

Svojstva fizikalnih raspada koji se dešavaju prilikom nastanka snopa čestica ne mogu se jednostavno očitati iz konačne distribucije čestica u detektoru. U ovom radu najprije će se simulirati sustav čestica jasno definiranih masa i kanala raspada koji predstavlja "istinit sustav" kojem želimo reproducirati distribucije vjerojatnosti raspada. Uz pretpostavku da imamo samo podatke koje sustav producira u detektoru, koristit ćemo metodu koja iterativno primjenjuje neuralnu mrežu u ulozi klasifikatora između događaja koje u detektoru ostavi istinita distribucija i događaja koje ostavi neka testna distribucija. Usporedit ćemo rezultate dobivene na ovaj način sa poznatim distribucijama raspada istinitog sustava. Rad je proširenje metode objavljene u Entropy 2020, 22(9), 994.

Abstract

The properties of decays that take place during jet formation cannot be easily deduced from the final distribution of particles in a detector. In this work, we first simulate a system of particles with well defined masses, decay channels, and probabilities. This is taken to be the “true system” for which we want to reproduce the decay probability distributions. Assuming we only have the data that this system produces in the detector, we decided to employ an iterative method which uses a neural network as a classifier between events produced in the detector by the “true system” and some arbitrary “test system”. In the end, we compare the distributions obtained with the iterative method to the “true” distributions. This work is an expansion of the method published in Entropy 2020, 22(9), 994.

Sadržaj

1 Uvod

Jedan od najčešćih sustava za istraživanje ponašanja elementarnih čestica visokoenergetski su sudari. Sudare proizvodimo u eksperimentima osmišljenim za otkrivanje i identifikaciju proizvedenih čestica. U načelu, od čestica koje se mogu proizvesti, samo su elektron, proton, foton i neprimjetni neutrini stabilni te se mogu detektirati. Sve ostale čestice se raspadaju. Međutim, čestice s vremenom života većim od $10^{-10}s$ nastale u sudarima visoke energije često posjeduju relativističke brzine pa mogu putovati i do nekoliko metara u detektoru i stoga se mogu izravno otkriti. U ovu klasu relativno dugovječnih čestica spadaju mioni, neutroni, nabijeni pioni i nabijeni kaoni. Kratko živuće čestice s vremenom života manjim od $10^{-10}s$ obično će se raspasti prije nego što proputuju značajnu udaljenost od mjesta proizvodnje i samo se njihovi proizvodi raspada mogu detektirati. Iz tog razloga potrebno je poznavati pravila njihovih raspada.

Posebnu kategoriju čestica nastalih u visokoenergetskim sudarima čine kvarkovi i gluoni. U sudarima visoke energije nastaju kvarkovi koji se zbog asimptotske slobode ponašaju kao kvazi-slobodne čestice, no zbog zatočenja boje, pojave koja uzrokuje da u prirodi nije uočen niti jedan slobodan kvark, njegovim udaljavanjem dolazi do hadronizacije. Hadronizacija je proces u kojoj udaljavanjem kvarkova nastaje kvark-antikvark par koji onda s inicijalnim kvarkovima stvaraju hadrone, čestice koje se sastoje od više kvarkova. Većina hadrona nema dugi vijek i oni se nastavljaju dalje raspadati na druge hadrone. Na kraju iz početnog kvarka određenog impulsa dobivamo snop čestica, čiji je impuls lokaliziran oko impulsa originalne čestice, koji nazivamo "jet". Distribucije nastanka i raspada pojedinih hadrona u procesu hadronizacije su nam nepoznate te će nam taj problem poslužiti kao motivacija za ovaj rad.

U drugom poglavlju ćemo se baviti kvantnom teorijom interakcije između kvarkova, kvantnom kromodinamikom. U njemu ćemo dodatno uvesti pojmove kao što su SU(3) baždarna invarijantnost, naboj boje, te gluoni, koji su prijenosnici sile između kvarkova. Zatim ćemo dodatno pojasniti pojave kao što su zatočenje boje i asimptotska sloboda. Na kraju ćemo dati kvalitativno objašnjenje same hadronizacije i proces stvaranja "jeta" od početnog kvarka. S obzirom da je cilj ovoga rada predložiti metodu iz koje iz skupa "jetova" pokušavamo doznati svojstva distribucija raspada

pojedinih čestica, predloženu metodu moramo testirati. U tu svrhu u trećem poglavlju konstruiramo tri različita načina generiranja "jetova" koje se međusobno razlikuju u svojstvima distribucija raspada. S obzirom da u njima pravila raspadanja znamo, to će nam poslužiti kao podloga koju želimo rekonstruirati i na kojoj možemo testirati učinkovitost predložene metode. Pošto je prisutna velika dimenzionalnost i kompleksnost problema i samog izračuna, u četvrtom poglavlju dati ćemo kratak osvrt na neuronske mreže koje su se pokazale pogodnim alatom u takvim situacijama. U petom poglavlju ćemo predstaviti samu metodu rekonstrukcije te izložiti tehničke detalje algoritma. Zatim ćemo navedeni algoritam primijeniti na podatke koje smo generirali po pravilima navedenim u trećem poglavlju. U šestom poglavlju ćemo napraviti analizu dobivenih rezultata našeg algoritma rekonstrukcije na tim podacima.

2 Kvantna kromodinamika (QCD)

Sredinom 20. stoljeća, uz pomoć sve snažnijih ubrzivača čestica, maglene ili Wilsonove komore, komore s mjehurićima te drugih mjernih instrumenata, otkriven je velik broj subatomskih čestica zvanih hadroni. Puno njih imalo je slična svojstva i mase, pa je postajalo jasno da sve ne mogu biti elementarne. Na temelju radova Murraya Gell-Manna i Georga Zweiga otkriveno je da se struktura hadrona može objasniti pomoću kombinacija manjih elementarnih čestica, kvarkova. Sila koja djeluje između kvarkova zove se jaka nuklearna sila te je jedna od četiri osnovne sile koje djeluju između elementarnih čestica uz elektromagnetsku, gravitacijsku i slabu nuklearnu silu. Detaljnom analizom pokazano je da se pojedini hadroni sastoje od kvarkova koji svi međusobno imaju ista do tada poznata kvantna stanja, što je dovelo do zaključka da kvarkovi imaju još jedan stupanj slobode, kasnije nazvan naboj boje. Kao što kod elektromagnetske sile naboj ima ulogu izvora, tako je naboj boje izvor jake nuklearne sile. Prenositelji sile su gluoni. U kvantnoj teoriji polja, kvantna kromodinamika (engl. *Quantum Chromodynamics* ili QCD) teorija je koja opisuje jaku interakciju između kvarkova i gluona te počiva na SU(3) baždarnoj invarijantnosti.

2.1 SU(3) baždarna invarijantnost, naboj boje i gluoni

Za razliku od kvantne elektrodinamike, koja počiva na U(1) baždarnoj simetriji, temeljna simetrija koja opisuje kvantnu kromodinamiku je SU(3) lokalna baždarna simetrija. Za danu valnu funkciju odnosno Diracov spinor $\psi(x)$, SU(3) transformacija lokalne faze može se zapisati kao [1]:

$$\psi(x) \rightarrow \psi'(x) = \mathbf{exp}[ig_s\, \boldsymbol{\alpha}(x)\, \hat{\mathbf{T}}]\psi(x)\,, \tag{2.1}$$

pri čemu je g_s faktor proporcionalnosti koji odgovara jakosti sile, a $\hat{\mathbf{T}}$ predstavlja osam generatora SU(3) grupe koji su povezani s Gell-Mannovim matricama. $\boldsymbol{\alpha}(x) = \{\alpha^a\}$ predstavlja osam funkcija prostorno-vremenske koordinate x. Zbog toga što su generatori SU(3) grupe 3x3 matrice, valna funkcija ψ mora sadržavati novi stupanj slobode koji možemo prikazati kao 3-komponentni vektor. Ovaj novi stupanj slobode

nazivamo boja, a moguća stanja označavamo crvenom (r), zelenom (g) i plavom (b):

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{i} \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.2}$$

Transformacija lokalne faze SU(3) odgovara "rotirajućim" stanjima u tom prostoru boja oko osi koja je različita u svakoj točki prostorvremena. Ako želimo invarijantnost na transformaciju 2.1, Diracova jednadžba postaje:

$$i\gamma^{\mu}[\partial_{\mu} + ig_s A^a_{\mu} T^a]\psi - m\psi = 0\,, \tag{2.3}$$

pri čemu su A^a_{μ} osam novih baždarnih polja koja odgovaraju generatorima SU(3) transformacije (a = 1, . . . , 8). Tih osam polja nazivamo gluonima i oni su prijenosnici QCD sile između kvarkova koji nose naboj boje. Slično kao što antičestice nose naboj suprotan česticama u QED-u, u kvantnoj kromodinamici antičestice nose naboje anti-boje $(\bar{r}, \bar{g}, \bar{b})$.
Feynmanovo pravilo za QCD vrh glasi [1]:

$$-\frac{1}{2} i g_s \lambda^{\mu}_{ij} \gamma^{\mu}\,, \tag{2.4}$$

pri čemu su λ^{μ}_{ij} Gell-Mannove matrice te i i j označavaju boje kvarkova. Slijedom toga, gluoni koji odgovaraju nedijagonalnim članovima Gell-Mannovih matrica povezuju stanja kvarkova različitih boja. Kako bi se boja sačuvala tijekom interakcije gluoni moraju nositi naboj boje (detaljnijom analizom dobiva se da gluoni moraju nositi istodobno i naboj boje i anti-bojni naboj). Nadalje, Feynmanovo pravilo za gluonski propagator glasi [1]:

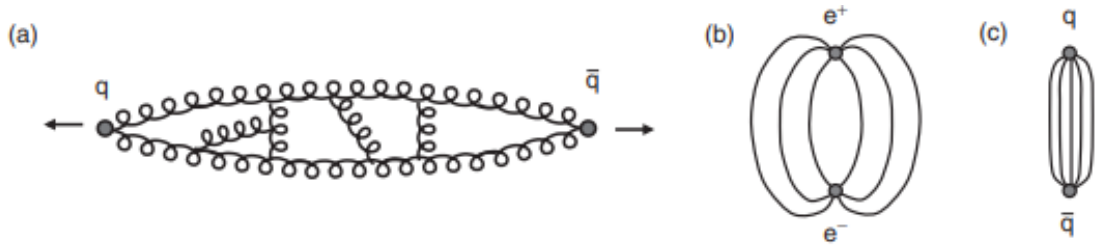
$$-i\frac{g_{\mu\nu}}{q^2}\delta^{ab}\,, \tag{2.5}$$

pri čemu delta funkcija osigurava da je gluon vrste a, koji se emitira na vrhu označenim s μ, jednaka onom koji se apsorbira na vrhu označenom s ν.

2.2 Zatočenje boje

Hipoteza zatočenja u QCD-u govori nam da su sva izolirana kvantna stanja bojno neutralna. Upravo zbog toga kvarkove, koji nose naboj boje, nikad ne nalazimo kao

slobodne čestice, već uvijek zatvorene unutar hadrona. Unatoč tome što ne postoji analitički dokaz za zatočenje boje, svi eksperimenti ukazuju na njezinu točnost. Ipak, kvalitativno razumijevanje vjerojatnog podrijetla hipoteze može se dobiti razmatranjem onoga što se događa kad se razdvoje dva kvarka. Interakcije između kvarkova mogu se promotriti u smislu razmjene virtualnih gluona. Za razliku od fotona koji ne nose nikakav naboj, gluoni nose naboj boje te zbog toga postoji privlačna sila između njih (Slika 2.1-a). Ta interakcija stišće polje boje između kvarkova (Slika 2.1c) u tzv. cijevi umjesto da se silnice polja šire kao u QED-u (Slika 2.1b).



Slika 2.1: Kvalitativna slika učinka interakcija gluona i gluona na dalekodosežnu QCD silu [1].

Na velikim udaljenostima, gustoća energije u cijevi je otprilike konstanta, što znači da potencijal poprima oblik:

$$V(r) \sim \kappa r \,, \tag{2.6}$$

pri čemu je r udaljenost između kvarkova. S obzirom da se energija pohranjena u polje boje linearno povećava s udaljenosti, razdvajanje dva kvarka zahtijeva beskonačno mnogo energije. Kad bi postojala dva slobodna naboja boje, energija njihovog polja bila bi beskonačna. Kao rezultat toga, obojeni objekti vezani su u hadronska stanja bez polja boja između njih. Slijedom toga, kvarkovi su uvijek ograničeni unutar bezbojnih hadrona.

2.3 Asimptotska sloboda

Iz renormalizacije konstante vezanja u QCD-u slijedi njezina ovisnost o energiji [2]:

$$\alpha(q^2) = \frac{\alpha(\mu^2)}{1 - \beta_0 \frac{\alpha(\mu^2)}{4\pi} \ln\left(\frac{q^2}{\mu^2}\right)} \,, \tag{2.7}$$

pri čemu su β_0 i $\alpha(\mu^2)$ konstanta koja ovisi o broju kvarkova i boja i znana vrijednost konstante vezanja pri impulsu μ. Možemo primijetiti da snaga QCD vezanja

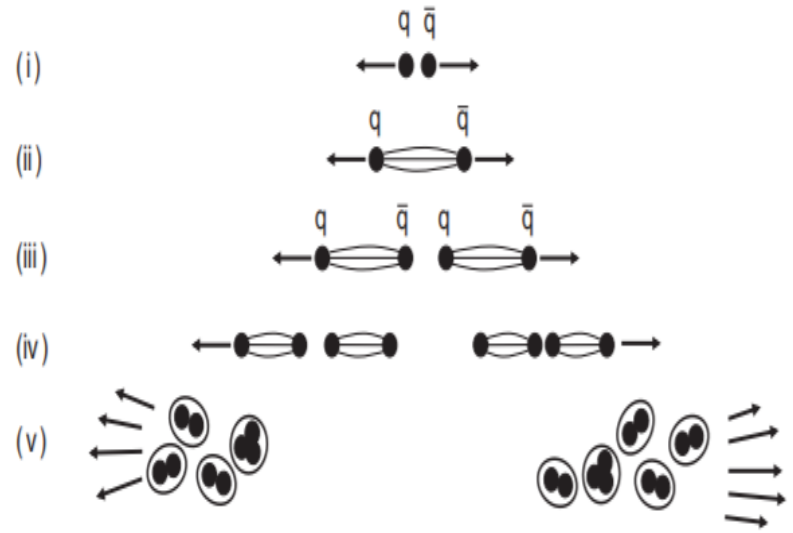
značajno varira u intervalu energija relevantnom za čestičnu fiziku. Konstanta vezanja u kvantnoj kromodinamici, za razliku kvantne elektrodinamike, velika je pri malim energijama. Samim time, perturbativni račun u ovom slučaju nije moguć. Na izmjenama imupulsa većim od $|q| > 100GeV/$c, što odgovara tipičnim vrijednostima u modernim visoko-energetskim eksperimentima na sudarivačima, konstanta vezanja α_s približno je jednaka 0.1, što je dovoljno malen iznos da izračunima možemo pristupiti perturbativno. Ovo svojstvo kvantne kromodinamike nazivamo asimptotska sloboda i razlog je zašto u pojedinim procesima kvarkove možemo promatrati kao kvazislobodne čestice.

2.4 Hadronizacija

Kvarkovi su dosad eksperimentalno opaženi samo kao zatočeni objekti. Međutim, u sudarima visokih energija, primjerice u procesu $e^+e^- \rightarrow q\overline{q}$, nastaju kvarkovi, a ne hadroni. Njihovim udaljavanjem, zbog zatočenja boje, u jednom trenutku postane energetski povoljniji nastanak novog kvark-antikvark para. Kao posljedica takvih interakcija, energija u polju jake interakcije dva kvarka pretvara se u još parova kvarkova i antikvarkova kroz proces hadronizacije. Proces hadronizacije odvija se na putu duljine oko 10^{-15} m te rezultira **snopom**, odnosno ***jetom*** hadrona. Dakle, pojedini kvark opažamo kao *jet* koji se uglavnom sastoji od nabijenih čestica i fotona.

Slika 2.2: Prikaz *jeta* u detektoru [1].

Trenutno je razumijevanje detalja procesa hadronizacije poprilično oskudno, ali postoji mnogo fenomenoloških modela koje eksperimentalni podaci podržavaju. Pet glavnih stadija hadronizacije prikazano je na Slici 2.3:



Slika 2.3: Pet glavnih stadija hadronizacije kvark – antikvark para $q\overline{q}$ [1].

(i) U interakciji nastane par kvark i antikvark $q\overline{q}$.

(ii) Tijekom njihova razdvajanja polje boja sadržano je u cijevi gustoće energije oko 1 GeV/fm.

(iii) Kako se kvarkovi nastavljaju odvajati, energija sadržana u polju boja dovoljna je za stvaranje novih $q\overline{q}$ parova te je razdvajanje polja boja u manje "cjevčice" energetski povoljnije.

(iv) Isti proces se nastavlja i proizvodi se još parova $q\overline{q}$ sve dok

(v) svi kvarkovi i antikvarkovi imaju dovoljno nisku energiju i spajaju se u bezbojne hadrone. Rezultat ovog procesa su dva "jeta" hadrona, jedan usmjeren u početnom smjeru kvarka te drugi usmjeren u početnom smjeru antikvarka [1].

3 Opis simuliranog sustava

U ovom poglavlju pobliže opisujemo izmišljene sustave čije ćemo ponašanje pokušati rekonstruirati. Bit će prisutna tri različita načina generiranja *jetova*. U svakom od njih krećemo od početne čestice poznate mase koja se iterativno raspada kroz tri iteracije, no distribucije pojedinih raspada će se razlikovati u tri slučaja. Najprije razmatramo postojanje samo diskretnog broja mogućih čestica te diskretnog broja načina na koje se pojedina čestica može raspasti. U drugom slučaju definirati ćemo sustav u kojem će biti prisutne kontinuirane distribucije raspada koje ćemo zatim koristiti kao pozadinu za treći sustav u kojem ćemo na pozadinu dodati nekoliko skrivenih rezonanca, tj. čestica s većim mogućnostima nastanka. *Jetove* generirane pomoću trećeg sustava koristiti ćemo za analizu razlučivosti samog algoritma rekonstrukcije.

3.1 Diskretni raspadi

Sustav u kojem promatramo diskretne raspade zamišljamo na način da imamo 10 vrsta čestica, označenih slovima od A do J, koje se međusobno raspadaju samo jedne u druge. Takav sustav zatvoren je s obzirom na ostale čestice izvan samog sustava. Mase pojedinih čestica (u proizvoljnim jedinicama) te njihove vjerojatnosti raspada na druge čestice prikazani su na slici 3.1. Pod pojmom "vjerojatnosti raspada" mislimo na odnose amplituda pojedinih raspada.

	A		B		C		D		E	
masa	0.1		0.6		1.3		1.9		4.4	
raspadi	1	A	0.7	B	1	C	0.3	A C	0.6	C C
			0.3	A A			0.3	A A	0.4	E
							0.4	D		

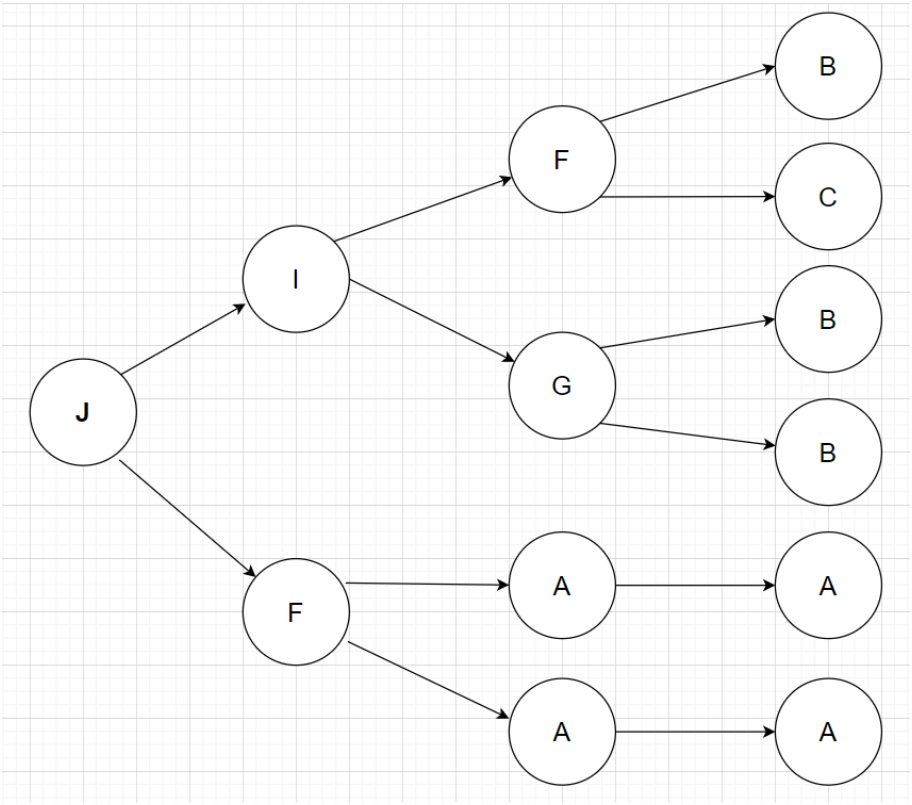
	F		G		H		I		J	
masa	6.1		8.4		14.2		18.1		25	
raspadi	0.5	A A	0.9	B B	0.6	D D	1	F G	0.5	F I
	0.5	B C	0.1	A F	0.25	D E			0.4	G H
					0.15	E F			0.1	E E

Slika 3.1: Tablica čestica u simuliranom sustavu.

Čestice A–E mogu se detektirati u zamišljenom detektoru jer su očekivani put koji prijeđu za svog života i veličina detektora istog reda veličine. U tablici se to može

očitati kao vjerojatnost da će se pojedina čestica “raspasti na sebe samu”, što zapravo znači da je njezino vrijeme života dovoljno dugačko da bi se mogla detektirati u tom stanju. Od navedenih, čestice A i C potpuno su stabilne.

Čestice F–I skrivene su rezonance: njihov očekivani životni vijek vrlo je malen te se one nikada neće detektirati. *Jet* nastaje raspadanjem početne čestice *J*, takozvane “majke”. Nakon toga slijede dvije dodatne iteracije raspada dobivenih čestica po već navedenim pravilima raspadanja. U svakom raspadu pretpostavljamo da je uniformna distribucija raspada na određene kutove. Na kraju dobivamo skup od maksimalno 8 čestica (A–E) sa zadanim impulsima u sustavu mirovanja početne čestice. Primjer jednoga takvog dobivenog *jeta* vidimo na Slici 3.2.



Slika 3.2: Primjer jednog raspada s navedenim pravilima.

3.2 Kontinuirane distribucije raspada (pozadina)

Početna čestica ($m = 25$), za razliku od prethodnog poglavlja, raspada se tako da je svaka kombinacija dvije mase koja zadovoljava fizikalne zakone očuvanja poslije tog raspada moguća. Također, pretpostavit ćemo da distribucija raspada ovisi “majci”, tj.

masi čestice koja se raspada. Raspad ćemo opisivati pomoću parametara *a* i *b*:

$$\begin{aligned} \frac{abM}{2} &= m_1\,, \\ aM\left(1-\frac{b}{2}\right) &= m_2\,, \end{aligned} \tag{3.1}$$

pri čemu parametar *a* predstavlja udio zbroja masa nastalih čestica u ukupnoj masi, dok parametar *b* predstavlja njihov omjer. Možemo primijetiti da ovakvom parametrizacijom vrijedi $m1 \leq m2$ te da je prostor koji pokrivaju ovakvo definirani parametri većinom zauzet raspadima u kojima nastaju manje čestice. To odgovara nekoj realnoj situaciji u kojoj očekujemo da se u produktima raspada nalaze manje mase zbog potisnuća faznim prostorom. Sada gradimo generalnu gustoću vjerojatnosti parametara u ovisnosti o ulaznoj masi. Počinjemo od:

$$\rho^0_{m,a}(a) = \frac{1}{\mathcal{N}_0(m)}\mathbf{exp}\left(\frac{-2000\cdot(a-0.75)^2}{2+m}\right), \quad \rho^0_{m,a}(b) = 1\,, \tag{3.2}$$

pri čemu je $\mathcal{N}_0$ normalizacijska konstanta i svi su parametri pravilno dimenzijski skalirani. S obzirom na to da je gustoća vjerojatnosti vezana uz parametre a i b, $\mathcal{N}_0$ je funkcija početne mase m. Za svaku ulaznu masu distribucija je različita i normirana.

Vidimo da je gustoća vjerojatnosti za parametar a normalna distribucija kojoj je srednja vrijednost 0.75, a devijacija joj je funkcija mase. Gustoća vjerojatnosti za parametar b je uniformna. Nakon toga dodajemo čestice koje se ne raspadaju ili imaju manju šansu za samo raspadanje. Definiramo da se raspad nije dogodio kad je parametar $a = 1$ i $b = 0$ jer je tad masa prve čestice u raspadu jednaka nuli, a masa druge čestice jednaka početnoj, pa iz toga slijedi da prva čestica nema ni energiju ni impuls. Stoga, ne postoji. Neka se čestice, čije su mase u okolini $m_i = 0, 0.7, 1.4, 2.5, 5$ ne raspadaju. Definiramo novu gustoću vjerojatnosti:

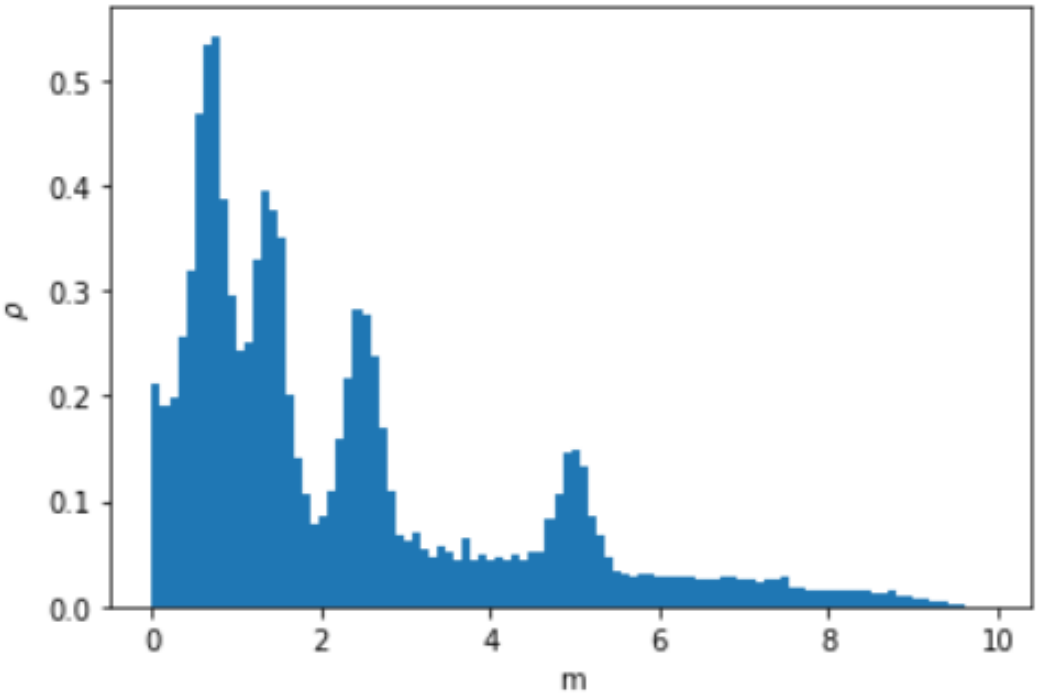
$$\begin{aligned} \rho^1_{m,a}(a) &= \frac{1}{\mathcal{N}^1_a(m)}\left(\sum_{i=1}^{5} C\cdot\mathbf{exp}\left(100\cdot(-1+a)\right)\cdot\mathbf{exp}\left(\frac{-(m-m_i)^2}{0.1}\right)+\rho^0_a(a)\right), \\ \rho^1_{m,b}(b) &= \frac{1}{\mathcal{N}^1_b(m)}\left(\sum_{i=1}^{5} C\cdot\mathbf{exp}(-100\cdot b)\cdot\mathbf{exp}\left(\frac{-(m-m_i)^2}{0.1}\right)+\rho^0_b(b)\right), \end{aligned} \tag{3.3}$$

čime s $C = 100$ omogućavamo da vjerojatnosti u okolinama navedenih početnih masa za $a = 1$ i $b = 0$ budu jako blizu jedan. $\mathcal{N}^1_a$ i $\mathcal{N}^1_b$ su nove normalizacijske konstante.

Na kraju još dodajemo dio koji zahtijeva da su raspadi u kojem su prisutne navedene mase kao produkti, vjerojatniji:

$$\begin{aligned} \rho_{m,p}(a,b) &= \sum_{i=1}^{5}\left(\mathbf{exp}\left(\frac{-(\frac{mab}{2}-m_i)^2}{0.7}\right)+\mathbf{exp}\left(\frac{-(am(1-\frac{b}{2})-m_i)^2}{0.7}\right)\right)\cdot C\,, \\ \rho_m(a,b) &= \frac{1}{\mathcal{N}(m)}\left(0.1\cdot(30-m)^2\cdot\rho_{m,p}(a,b)+\rho^1_{m,a}(a)\cdot\rho^1_{m,b}(b)\right), \end{aligned} \tag{3.4}$$

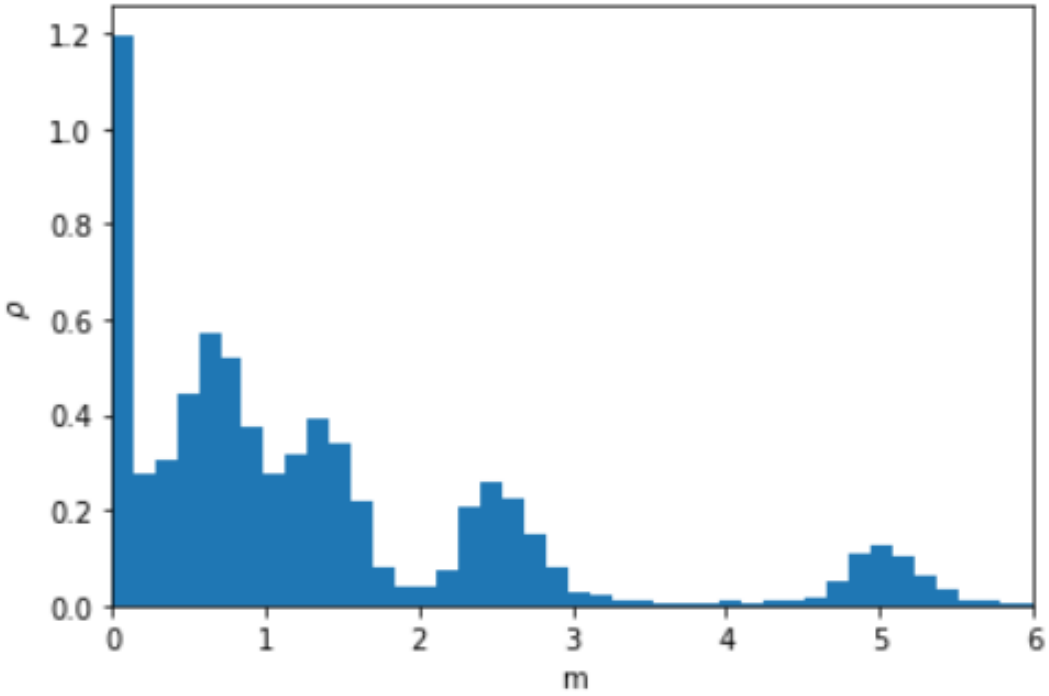
pri čemu je $C = 6$. Bitno je napomenuti da su prisutne konstante i funkcije, kao i interval samih masa koje koristimo u ovom sustavu, proizvoljne. Naglasak ovdje više stoji na kvalitativnom ponašanju samih distribucija. Također, cilj ovog rada leži u rekonstrukciji ovdje zadanih gustoća vjerojatnosti raspada. Na slici 3.3, koja prikazuje gustoću vjerojatnosti raspada čestice mase $m = 10$ kao funkciju masa, uočavamo vrhove na mjestima gdje se nalaze mase stabilnih čestica.



Slika 3.3: Distribucija nastalih masa u raspadu čestice mase $m = 10$.

Kao i u slučaju diskretnih raspada, u računu krećemo od početne čestice mase $m = 25$ u sustavu mirovanja koja se raspada na dvije nove čestice pomoću slučajnih varijabla a i b s gustoćama vjerojatnosti opisanim u relaciji (3.4). Nadalje, svaka se novonastala čestica raspada pomoću istih pravila. Taj postupak ponavljamo tri puta počevši od prve čestice. Na kraju raspolažemo s listom od maksimalno 8 čestica (nije nužno 8 jer može doći do ne-raspada čestice) s određenim energijama i impulsima

u sustavu mirovanja početne čestice. Slika 3.4 prikazuje histogram "detektiranih" čestica, tj. čestica koje su se pojavile nakon treće iteracije, nakon koje zaustavljamo postupak. Možemo primijetiti da se detektirane mase nalaze u okolini stabilnih čestica ($m_i = 0, 0.7, 1.4, 2.5, 5$). Ističemo još jednom razliku u distribucijama sa slika 3.3 i 3.4: na prvoj slici prikazana je distribucija masa koje nastaje u velikom broju raspada čestice određene mase (koja u prikazanom slučaju iznosi 10), dok je na drugoj slici prikazana distribucija masa čestica koje u detektoru ostavi čestica početne mase 10 nakon 3 iteracije raspada.



Slika 3.4: Distribucija masa "detektiranih" čestica.

3.3 Kontinuirani raspadi sa skrivenom rezonancom

Treći sustav koji će biti predmet analize u ovom radu samo je nadogradnja na sustav opisan u prošlom potpoglavlju. Način generiranja *jetova* biti će jednak,a gustoća vjerojatnosti sadržavati će dodatne članove iz relacije (3.4). Definirati ćemo dvije skrivene rezonance, tj. čestice s masom u blizini skrivene rezonance imati će veću vjerojatnost nastanka nego što je to bilo u prethodnom potpoglavlju. Također, skrivene rezonance neće imati veliku vjerojatnost ne-raspada i same detekcije, već će najveća vjerojatnost biti da se raspadnu na čestice masa $m_i = 1.4, 2.5, 5$ koje su bile opisane kao čestice koje se dalje ne raspadaju. Dvije skrivene rezonance imaju mase $m_j = 6.5, 12$.

Uvodimo aktivacijske funkcije $f_j^0(m)$ koje nam govore pri kojim je početnim masama raspad na pojedinu skrivenu rezonancu relevantan:

$$\begin{aligned} f_1^0(m) &= \mathbf{exp}\left(-\frac{(m-15)^2}{30}\right), \\ f_0^0(m) &= \mathbf{exp}\left(-\frac{(m-20)^2}{15}\right). \end{aligned} \tag{3.5}$$

Zatim definiramo funkciju $\rho^0_{m,hid}(a,b)$ koja opisuje raspade na skrivene rezonance:

$$\begin{aligned} \rho^0_{m,hid}(a,b) =& f_j^0(m) \cdot \sum_{j=1}^{2} \left(\mathbf{exp}\left(\frac{-(\frac{mab}{2} - m_j^{hid})^2}{0.5}\right) \right. \\ &\left. + \mathbf{exp}\left(\frac{-(am(1-\frac{b}{2}) - m_j^{hid})^2}{0.5}\right)\right). \end{aligned} \tag{3.6}$$

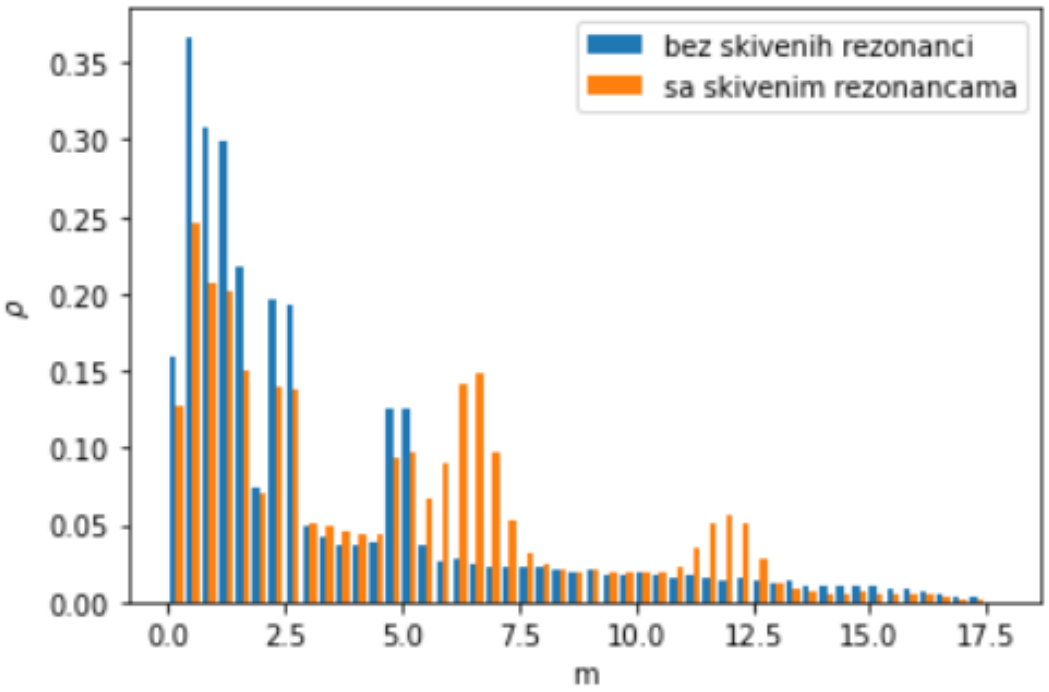
Nakon toga promatramo raspad skrivenih rezonanci. U ovom slučaju aktivacijska funkcija $f_j^1(m)$ poprima relevantne vrijednosti samo kad je m blizu masa skrivenih rezonanci m_j. U raspadu čestica mase $m_0 = 6.5$, produkti u blizini masa $m_{0,k} = 1.2, 2.5$ imaju veću šansu nastanka, a u raspadu čestice mase $m_1 = 12$ vjerojatniji su produkti čestice $m_{0,k} = 2.5, 5$. To je opisano u funkciji $\rho^1_{m,hid}(a,b)$:

$$\begin{aligned} f_j^1(m) =& \mathbf{exp}\left(-\frac{(M - m_j^{hid})^2}{0.2}\right), \\ \rho^1_{m,hid}(a,b) =& f_j^1(m) \cdot \\ & \sum_{j=1}^{2}\sum_{k=1}^{K_j} \left(\mathbf{exp}\left(\frac{-(\frac{mab}{2} - m_{j,k})^2}{0.5}\right) + \mathbf{exp}\left(\frac{-(am(1-\frac{b}{2}) - m_{j,k})^2}{0.5}\right)\right), \\ \rho_{m,hid}(a,b) =& \rho^1_{m,hid}(a,b) + \rho^0_{m,hid}(a,b). \end{aligned} \tag{3.7}$$

K_j je konstanta koja govori koliko skrivena rezonanca ima vjerojatnijih produkata. U našem slučaju, za obje skrivene rezonance ona je jednaka 2. Ukupan doprinos skrivenih rezonanci promatramo kao zbroj funkcije koja opisuje raspade na skrivene rezonance te same raspade skrivenih rezonanci. Na kraju gustoći vjerojatnosti definiranoj u relaciji (3.4) pridodajemo funkciju dobivenu u (3.7) te normiramo da dobijemo novu gustoću vjerojatnosti koju ćemo koristiti u generiranju *jetova*.

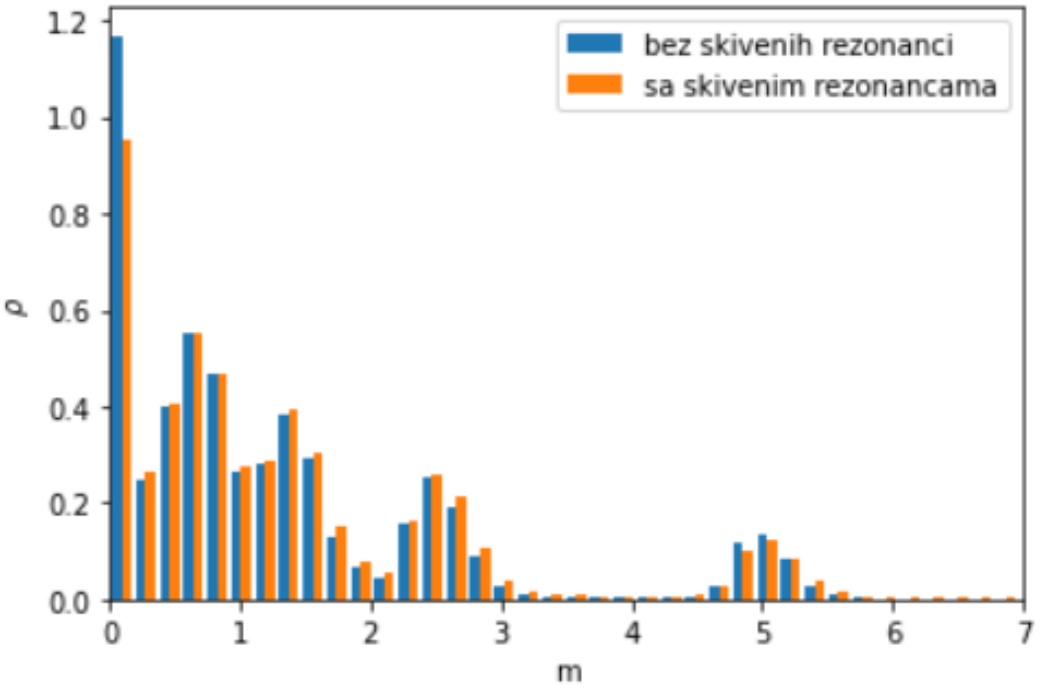
$$\rho_{m,res}(a,b) = \frac{1}{\mathcal{N}_{res}(m)}\Big(C \cdot \rho_{m,hid}(a,b) + \rho_m(a,b)\Big), \tag{3.8}$$

pri čemu smo koristili $C = 400$. Kao što je spomenuto, sam postupak generiranja *jetova* ostaje isti kao u prethodnom potpoglavlju te početna čestica opet ima masu 25. Na slici 3.5 vidimo razliku distribucija iz relacija (3.4) i (3.8) za česticu mase $m = 18$. Iz slike se jasno vidi vrh u okolini masa skrivenih rezonanca kojeg u prijašnjoj distribuciji nije bilo.



Slika 3.5: Primjer distribucije raspada čestice mase $m = 18$ za kontinuirane distribucije raspada bez skrivenih rezonanci (plavo) i za kontinuirane distribucije raspada sa skrivenim rezonancama (narančasto).

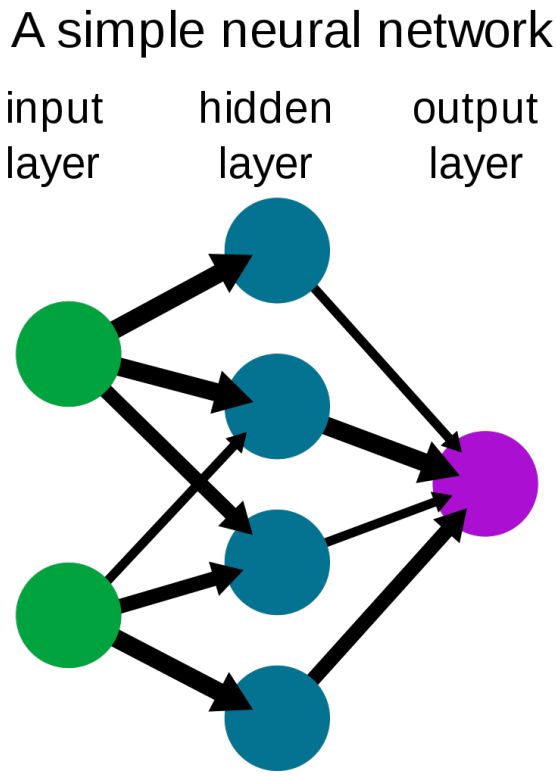
Na slici 3.6 prikazana je usporedba distribucija masa "detektiranih" čestica u generiranim *jetovima* po pravilima opisanim u prethodnom potpoglavlju i po pravilima opisanim u ovom potpoglavlju. Za razliku od slike 3.5 gdje se vidi jasna razlika između dvije distribucije, na slici 3.6 ne vidimo jasnu razliku između dvije distribucije. Distribucija dobivenih masa dosta je slična te se razlika između njih krije u drugim informacijama. Iz tog razloga ovaj primjer koristimo da procijenimo koliko će naša metoda rekonstrukcije (opisana u 5. poglavlju) uspjeti uhvatiti postojanje skrivenih rezonanca.



Slika 3.6: Usporedba distribucije masa “detektiranih” čestica generiranim u *jetovima* generiranim po relaciji (3.4), označeno plavom bojom te generiranim po relaciji (3.8), označeno narančastom bojom.

4 Neuronske mreže

Neuronske mreže jedna su od najraširenijih metoda strojnog učenja. Osnovni koncept mreže osmišljen je sredinom 20. stoljeća te je od tada zanimljivo područje istraživanja, kako u znanosti, tako i u industriji.



Slika 4.1: Primjer jednostavne neuronske mreže. Preuzeto san [3]

Struktura mreže inspirirana je strukturom ljudskog mozga te se sastoji od nekoliko slojeva međusobno povezanih čvorova (neurona). Prvi sloj predstavlja ulaz mreže, dok zadnji sloj predstavlja njezin izlaz. U osnovnom je modelu vrijednost svakog neurona u nekom sloju jedinstvena kombinacija vrijednosti neurona iz prethodnog sloja. Formalnije, ako s $\mathbf{x}^k$ označimo $k-ti$ sloj mreže, a s x_i^k, $i = 1, \dots n_k$ vrijednosti neurona u tom sloju, tada je vrijednost j-tog neurona u sljedećem sloju jednaka

$$x_j^{k+1} = \sigma\Big(\sum_{i=1}^{n} w_{ij}^k x_i^k + b_j^k\Big), \tag{4.1}$$

pri čemu je σ aktivacijska funkcija. Potrebu za aktivacijskim funkcijama i najčešće korištene oblike aktivacijskih funkcija objasnit ćemo u potpoglavlju 4.2.

Ako primjenu funkcije σ na vektor shvatimo kao primjenu na svaki član vektora

pojedinačno, gornju relaciju možemo zapisati kao:

$$\mathbf{x}^{k+1} = \sigma\Big(W^k\mathbf{x}^k + \mathbf{b}^k\Big)\,.$$

Ovaj osnovni model može se poboljšati i prilagoditi mijenjanjem broja slojeva i neurona u pojedinom sloju, uvođenjem nekih restrikcija na prostor parametara te kombiniranjem jednostavnih mreža u kompliciranije strukture. Napreci u samom računanju, kao što su korištenje stohastičkog gradijentnog spusta i paralelna implementacija algoritama na grafičkim karticama, omogućili su da se i kompliciranije strukture mogu istrenirati u razumnom vremenu.

4.1 Učenje neuronske mreže

Učenje jednostavne neuronske mreže podrazumijeva poznavanje željenog izlaza iz mreže za velik broj podataka (tzv. označeni podaci). *Loss* funkcijom L opisujemo podudarnost izlaza $\hat{y}(x, \mathbf{W}, \mathbf{b})$ iz trenutne mreže i željenog izlaza y. Ako je cilj mreže regresija (procjena vrijednosti neke varijable), najčešće se koriste L^1 i L^2 norma greške, dane formulama:

$$L(\mathbf{W}, \mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i(x_i, \mathbf{W}, \mathbf{b}) - y_i|\ ,$$

$$L(\mathbf{W}, \mathbf{b}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{y}_i(x_i, \mathbf{W}, \mathbf{b}) - y_i\right)^2}\,.$$

U slučaju da je cilj mreže binarna klasifikacija (svrstavanje podatka u jednu od dvije kategorije), najčešće korištena *loss* funkcija je *binary cross entropy loss* [4]:

$$L(\mathbf{W}, \mathbf{b}) = -\Big(\sum_{i=1}^{n} t_i \log(\hat{y}_i(x_i, \mathbf{W}, \mathbf{b})) + (1 - t_i)\log(1 - \hat{y}_i(x_i, \mathbf{W}, \mathbf{b}))\Big)\,,$$

pri čemu je $t_i = 0$ ako i-ti podatak pripada prvoj grupi, a $t_i = 1$ ako pripada drugoj grupi. Izlaz iz mreže za svaki podatak tada poprima vrijednost u $[0, 1]$ i predstavlja predviđenu vjerojatnost da podatak pripada drugoj grupi. Gornja se formula može poopćiti i za klasifikaciju u M kategorija.

Učenje neuronske mreže zapravo je proces minimizacije *loss* funkcije L s obzirom na parametre $\mathbf{W}$ i $\mathbf{b}$. Ta se minimizacija u praksi najčešće radi pomoću gradijentnog spusta, iterativnog algoritma u kojem u sljedeće stanje dolazimo kretanjem u smjeru

suprotnom od gradijenta *loss* funkcije u trenutnoj točki. Formalno, ako s $\mathbf{W}_n$ označimo n-tu točku u pretraživanju prostora parametara $\mathbf{W}, \mathbf{b}$, sljedeća je točka:

$$\mathbf{W}_{n+1} = \mathbf{W}_n - \gamma \nabla L(\mathbf{W}_n)\,. \tag{4.2}$$

Vrijednost parametra γ određuje brzinu učenja mreže te ga je potrebno ispravno namjestiti: premala vrijednost uzrokuje dugotrajniji proces učenja, a zbog prevelike vrijednosti stvarni lokalni minimum može se uvijek "preskakati" što onemogućuje nalaženje stvarnog rješenja. Gradijent iz (4.2) određujemo metodom propagacije unatrag (eng. *backpropagation*).

4.2 Aktivacijske funkcije

Kad u jednadžbi (4.1) ne bi postojala funkcija σ ili kad bi ona bila linearna, vrijednosti neurona sljedećeg sloja bile bi samo linearna funkcija vrijednosti neurona iz prethodnog sloja. Kako je linearna funkcija linearne funkcije opet linearna funkcija, prijelaz iz sloja k u sloj $k+2$ mogao bi se ekvivalentno napraviti u samo jednom sloju. Cijela neuronska mreža tad bi postala obična linearna regresija što jasno pokazuje potrebu na nelinearnom aktivacijskom funkcijom.

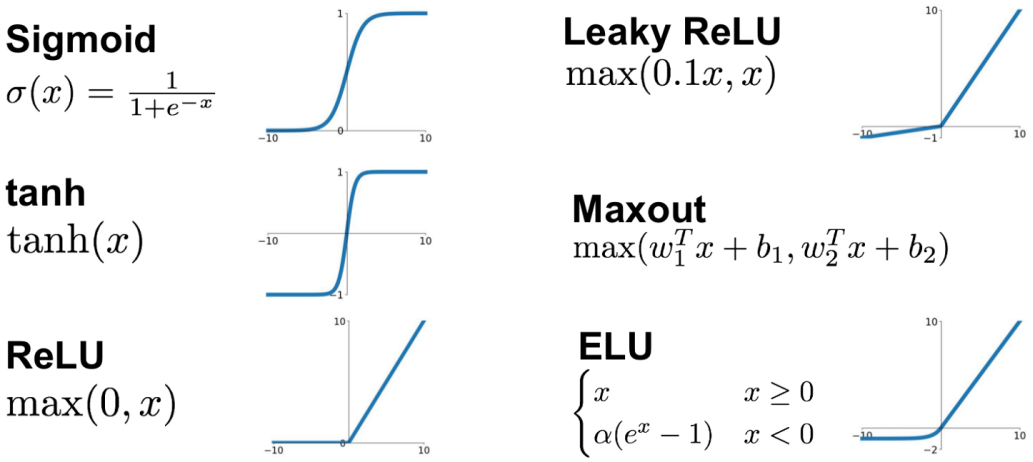
U originalno zamišljenom modelu neuronske mreže, funkcija σ bila je step funkcija, što je trebalo simulirati paljenje i gašenje neurona, kao u ljudskom mozgu. Međutim, zbog korištenja gradijentnog spusta, potrebno je koristiti derivabilne, ili barem neprekidne, funkcije. Iz tog razloga, na popularnosti je dobila funkcija sigmoid, odnosno funkcija oblika:

$$\sigma(x) = \frac{1}{1+e^{-x}}\,.$$

Sigmoid nalikuje na "zaglađenu" step funkciju te je i dalje često korišten, pogotovo u zadnjem sloju klasifikatora. Međutim, derivacija sigmoida je odozgo ograničena s $\frac{1}{4}$, što u dubokim mrežama uzrokuje tzv. *problem iščezavajućeg gradijenta* i onemogućuje učenje mreže.

U posljednje je vrijeme najpopularnija aktivacijska funkcija *ReLU*: funkcija koja je 0 za $x \leq 0$, a jednaka x za $x \geq 0$. U malom broju slučajeva može se dogoditi da dio jednak nuli uzrokuje da neki neuroni postanu beskorisni, što je motiviralo korištenje *Leaky ReLU* funkcije, gdje je dio jednak 0 zamijenjen linearnom funkcijom jako sporog

rasta te *ELU*, gdje je isti zamijenjen eksponencijalnom funkcijom.



Slika 4.2: Najčešće korištene aktivacijske funkcije. Preuzeto sa [5]

4.3 Konvolucijske neuronske mreže

Konvolucijske neuronske mreže posebna su klasa neuronskih mreža koja se najčešće koristi kad je ulazni podatak slika. Na slikama je većina značajki lokalnog karaktera, a konvolucijski slojevi to iskorištavaju na način koji je pogodan za deriviranje.
Za dvije slike a i b definiramo konvoluciju kao:

$$(a * b)[i, j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a[m, n]b[i - m, j - n] \,.$$

Iako su u ovom izrazu sume beskonačne, konačne slike zapravo možemo promatrati kao beskonačne ako postavimo vrijednosti 0 izvan granica prave slike.

Konvolucijski sloj u mreži zapravo je konvolucija ulazne slike s matricom manjih dimenzija od ulaza koju zovemo *filter*. Pojednostavljeno rečeno, manja matrica šeta po slici i djeluje na komadiće slike tih dimenzija. Mreža tada uči parametre u matrici filtra, a ulazno se mogu zadavati parametri poput veličine filtra, razmaka (koji "prorjeđuje" konvoluciju i zapravo skalira sliku) i nadopuna (kojom možemo kontrolirati dimenziju izlazne slike. Također možemo zadati i sami broj takvih filtera. Konvolucijski sloj ne mora nužno biti dvodimenzionalan, tj može se poopćiti za bilo koju dimenziju ($N \geq 1$), međutim princip ostaje isti.

5 Metoda rekonstrukcije

U poglavlju 3 definirali smo par načina nastanka *jetova*. Od sada nadalje uzimamo da su podaci proizvedeni po navedenim pravilima rezultati eksperimenta u kojem ne znamo fizikalnu pozadinu, tj. u ovom slučaju ponašamo se kao da su nam distribucije raspada pojedinih čestica nepoznate te ih želimo otkriti. U ovom poglavlju predlažemo metodu za rekonstrukciju navedenih parametara pomoću "testnih" distribucija i rezultata klasifikatora koji uči pripada li pojedini jet "testnom" skupu, izgeneriranom pomoću proizvoljnih testnih distribucija, ili pripada takozvanom "pravom" skupu podataka koji su napravljeni pomoću pravila koje želimo otkriti.

U prvom dijelu ovog poglavlja opisat ćemo samu srž metode pomoću jednog jednostavnog primjera. Zatim ćemo u sljedećim potpoglavljima dati uvid u pojedine dijelove algoritma, koje ćemo na kraju spojiti u cjelinu.

5.1 Jednostavan primjer

U ovom odjeljku pokušati ćemo pojasniti ideju korištenja klasifikatora za rekonstrukciju neke distribucije. Klasifikator treniran na dva skupa podataka kao izlaz daje vrijednost (engl. *score*) $s \in [0,1]$, pri čemu *score* 0 označava sigurnu pripadnost "testnom" skupu, a *score* 1 sigurnu pripadnost "pravom" skupu. Za podatak $x \in \mathbb{R}^m$, dvije distribucije ρ_{pravi}, $\rho_{\text{test}} \in \mathbb{R}^k$, vrijednost koju daje klasifikator, koju označavamo s $C_{\text{NN}}(x)$, u idealnoj je situaciji jednaka:

$$C_{\text{NN}}(x) = \frac{\mathcal{L}(\rho_{\text{pravi}}|x)}{\mathcal{L}(\rho_{\text{pravi}}|x) + \mathcal{L}(\rho_{\text{test}}|x)}\,, \tag{5.1}$$

pri čemu je $\mathcal{L}(\Theta_{\text{i}})$ vjerojatnost (engl. *likelihood*) da podatak x pripada distribuciji Θ_{i}. Nadalje, ako parametri distribucija Θ određuju neprekidnu vjerojatnosnu distribuciju, tad vrijedi:

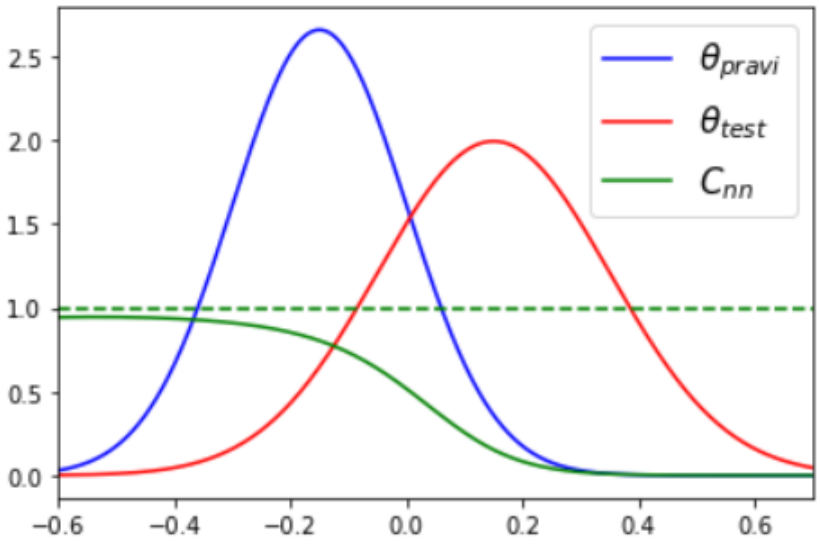
$$\begin{aligned} \mathcal{L}(\rho_{\text{pravi}}|x) &= \rho_{\text{pravi}}(x)\,, \\ \mathcal{L}(\rho_{\text{test}}|x) &= \rho_{\text{test}}(x)\,, \end{aligned} \tag{5.2}$$

pri čemu su $\rho(x|\Theta_{\text{i}})$ gustoće vjerojatnosti pripadnosti podatka x nekoj distribuciji Θ_{i}. Inverzijom izraza (5.1) sad možemo eksplicitno odrediti $\rho(x|\Theta_{\text{pravi}})$ uz poznate

$\rho(x|\Theta_{\text{test}})$ i izlaze klasifikatora C_{NN}:

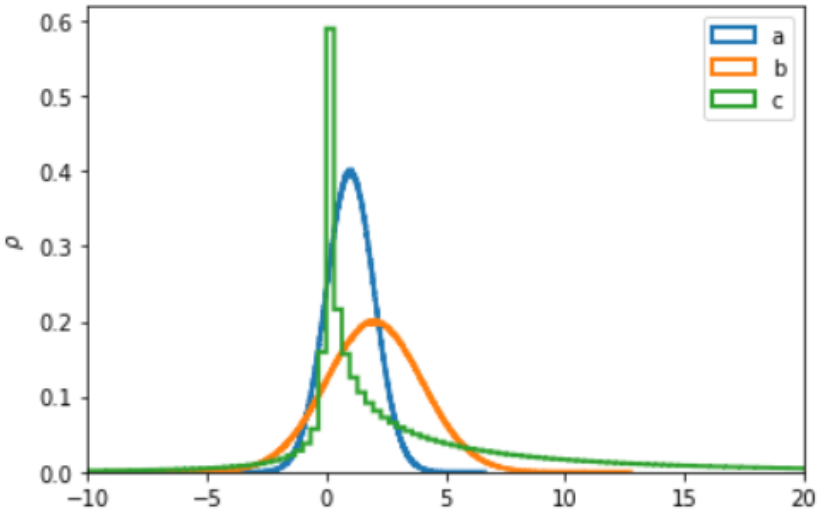
$$\rho_{\text{pravi}}(x) = \frac{C_{\text{NN}}(x)}{1 - C_{\text{NN}}(x)} \rho_{\text{test}}(x)\,. \tag{5.3}$$

Na slici 5.1 možemo vidjeti odnos distribucija vjerojatnosti i izlaza iz klasifikatora na primjeru dvije normalne razdiobe.



Slika 5.1: Primjer gustoća i rezultata pripadajućeg klasifikatora.

Promotrimo sad slučajnu varijablu c koja se dobiva umnoškom dvije nezavisne slučajne varijable a^2 i b. Na slici 5.2 vidimo primjer jedne takve distribucije gdje su a i b normalne raspodjele i dobivena je raspodjela slučajne varijable c.



Slika 5.2: Gustoća vjerojatnosti slučajne varijable a (normalna raspodjela sa sredinom 1 i devijacijom 1) i b (normalna raspodjela sa sredinom 2 i devijacijom 2) te slučajna varijabla c koja je definirana s $c = a^2 b$.

Ono što eksperiment mjeri samo je slučajna varijabla *c* dok *a* i *b* ostaju nepoznate. Za njihov pronalazak predlažemo sljedeća tri koraka:

1. Konstrukcija **generatora**. U ovom slučaju konstrukcija generatora je jednostavna. Pretpostavimo testne distribucije $\rho_{test,a}$ i $\rho_{test,b}$ za slučajne varijable *a* i *b* te zatim generiramo $\rho_{test,c} = \rho^2_{test,a} \cdot \rho_{test,b}$. Potrebno je pažljivo odabrati testne distribucije na način da je interval neiščezavajućih vrijednosti testnih distribucija nadskup intervala neiščezavajućih vrijednosti pravih distribucija.

2. Učenje **klasifikatora**. Odabiremo N podataka iz skupa eksperimentalno dobivenih vrijednosti za varijablu *c*, tj. iz distribucije $\rho_{pravi,c}$ te isto tako odaberemo N podataka iz distribucije $\rho_{test,b}$, tj. iz naših generiranih podataka. Klasifikator treniramo da nauči iz kojeg skupa podatak x dolazi. Podatke iz prave distribucije (eksperimenta) označavamo s 1, dok generirane podatke iz testne distribucije označavamo s 0.

3. **Izračunavanje gustoća vjerojatnosti**. Invertiranjem jednadžbe (5.1) možemo dobiti omjer gustoća vjerojatnosti $\rho_{pravi,c}(x)$ i $\rho_{test,c}(x)$ za podatak x. Znamo da vrijedi $\rho_{test,c}(x) = \rho^2_{test,a}(x_a) \cdot \rho_{test,b}(x_b)$, gdje je $x = x_a^2 \cdot x_b$. Uzimajući neku konstantnu vrijednost $x_b = B$ i računajući:

$$\rho_{pravi,a}(x_a)^2 \cdot \rho_{pravi,b}(B) = \rho_{test,a}(x_a)^2 \cdot \rho_{test,b}(B) \cdot \frac{\rho_{pravi,c}(x_a^2 \cdot B)}{\rho_{test,c}(x_a^2 \cdot B)} \tag{5.4}$$

   za svaki x_a, s obzirom ma to da poznajemo svaki član na desnoj strani, normiranjem možemo dobiti distribuciju $\rho_{pravi,a}$. Možemo primijetiti da će normiranjem, član $\rho_{test,b}(B)$ nestati jer je konstanta [6]. Računanje prave distribucije slučajne varijable *b* analogno je (tada stavljamo $x_a = A$ kao konstantni član). U ovom slučaju zbog jednostavnosti problema jednadžbu (5.4) moguće je napisati i riješiti egzaktno.

U našem slučaju eksperimentalni podatak x kojim raspolažemo bit će lista čestica u *jetu* dok će generator i samo računanje distribucija biti mnogo kompleksniji. Iako smo ovdje ilustrirali način na koji izlaz klasifikatora i nama poznate testne distribucije možemo koristiti za rekonstrukciju prave razdiobe, u praksi, susrećemo se s nekoliko problema: vrlo visoka dimenzionalnost prostora parametara i međuzavisnost dozvo-

ljenih procesa (zbog ovisnosti distribucije o masi) onemogućuju eksplicitno rješavanje problema te za procjenu prave distribucije koristimo neuronsku mrežu.

5.2 Generator

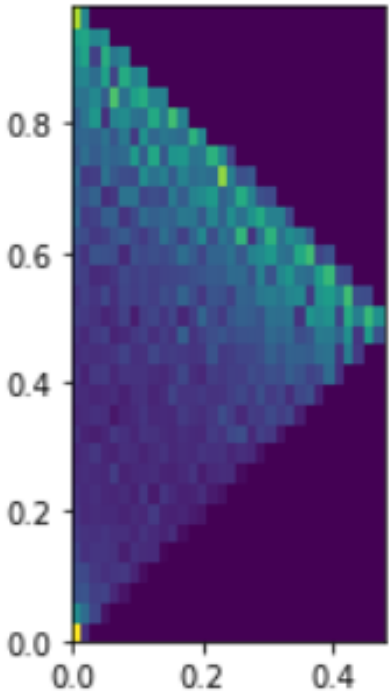
U ovom odjeljku dajemo opis generiranja *jetova* pomoću naših "testnih" distribucija.

1. Počinjemo od čestice u mirovanju mase $m_0 = 25$.

2. Čestica se dijeli na dvije nove čestice. Mase i momente novih čestica određujemo na sljedeći način:

   (a) Promatramo familiju vjerojatnosnih gustoća $\rho_M : \mathbb{R}^2 \to \mathbb{R}$, $M \in \mathcal{M}$. Treba naglasiti da je u teoriji skup brojeva koji mogu biti masa neke čestice konačan, ali nije unaprijed poznato koji su to brojevi.

   (b) Za dani M iz distribucije određene s ρ_M određujemo par (a, b) uz $a \in [0, 0.5]$, $b \in [0, 1]$ te ograničenja $a < b$ i $a + b \leq 1$. Gustoće ρ_M modeliramo prema dijelovima konstantne funkcije na mreži dobivenoj dijeljenjem intervala $[0, 1]$ na 50 podintervala. Primijetimo da će od mreže (25 x 50) preostati samo 650 mogućih parova *a* i *b*. Primjer toga je dan na slici 5.3.



Slika 5.3: Primjer distribucije parametara *a* i *b*.

Sad računamo:

$$\begin{aligned} aM &= m_1\,, \quad bM = m_2\,, \\ E_1 &= \left(\frac{M^2 + m_1^2 - m_2^2}{2M}\right)\,, \\ p_1 &= \sqrt{E_1^2 - m_1^2}\,, \\ p_2 &= -p_1\,, \quad E_2 = M - E_1\,. \end{aligned} \tag{5.5}$$

U slučaju da se čestica mase 0 raspada, tada su E_1 i p_1 jednaki 0, tj. ona iščezava dok druga čestica poprima isti impuls i energiju od početne čestice. Također, kad je parametar $a = 0$ te $b = 1$ ne dolazi do raspada. E_1, p_1, m_1 su tada opet 0, a impuls druge čestice je ponovo jednak impulsu početne čestice.

(c) Time smo odredili mase i količine gibanja novonastalih čsetica. Njihovu prostornu distribuciju postavljamo kao uniformnu u kutovima (θ, ϕ), $\theta \in [0, \pi], \phi \in [0, 2\pi]$.

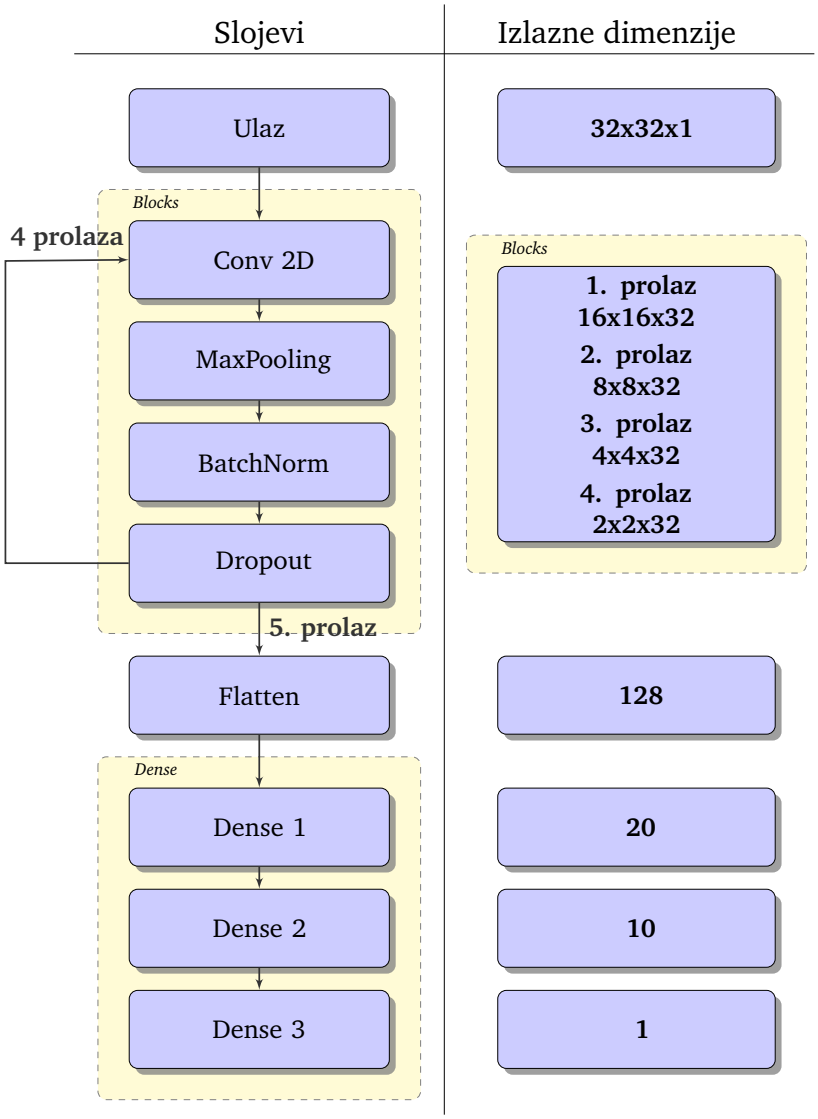
3. Nakon prvog raspada, proces 2 ponavljamo za svaku od dvije dobivene čestice 3 puta, nakon čega dobivamo najviše 8 čestica. Situacija gdje se čestica nije nastavila dijeliti odgovara odabiru parametra $a = 0$. Treba naglasiti da prilikom ponavljanja procesa za svaku novu česticu parametre a i b biramo iz druge distribucije jer su distribucije ovisne o masi. Dodatno, parametri za pojedinu distribuciju odnose se na sustav centra mase nove čestice pa ih treba i Lorentz-transformirati u sustav centra mase početne čestice, koji se poklapa s referentnim laboratorijskim sustavom.

4. Konačnu listu podataka s energijama i količinama gibanja svake od n čestica nazivamo *jet*. *Jet* je određen sa $1 + 2 + 4 = 7$ parova (a, b) te 7 parova (θ, ϕ), tj. ukupno 28 parametara.

5.3 Klasifikator

Klasifikator koji koristimo za pronalaženje prave familije distribucija jest *feed forward* konvolucijska neuronska mreža (engl. *convolutional neural network* ili CNN). Korištena arhitektura prikazana je na slici 5.4. Sastoji se od bloka slojeva u kojemu se nalaze dvo-

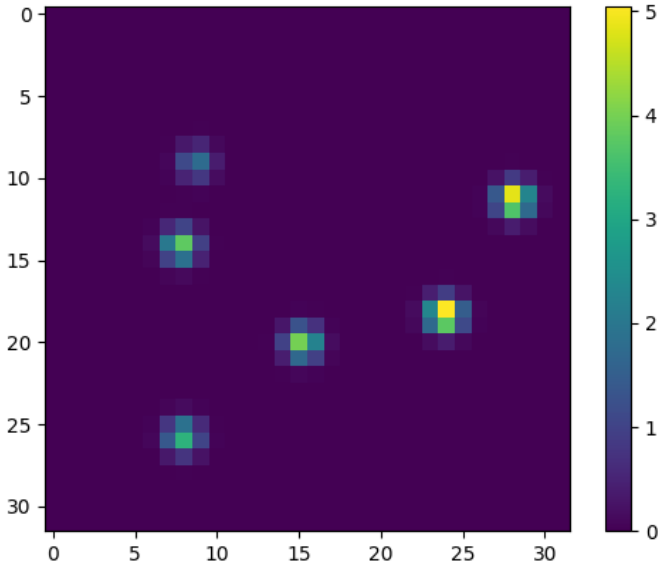
dimenzionalni konvolucijski sloj (s 32 filtera i (3,3) jezgrom ili kernelom), *MaxPooling* sloja te *batch normalization* i *dropout* sloja. Taj blok ponavlja se 4 puta. Nakon njega slijedi *flatten* sloj te 3 *dense* sloja, gdje je u prva dva korištena ispravljena linearna aktivacijska funkcija (engl. *rectified linear activation function* ili ReLU), dok je na zadnjem korištena sigmoid aktivacijska funkcija. Klasifikator treniramo tako da minimiziramo *binary cross entropy loss* [4], a za optimiziranje težina same konvolucijske neuronske mreže koristimo optimizator AdaM [7].



Slika 5.4: Arhitektura klasifikatora.

Jet vizualno prikazujemo histogramom (slikom) dimenzija 32x32 piksela. Osi histograma predstavljaju kutove (θ, ϕ) diskretizirane na 32 moguće vrijednosti, dok je boja svakog piksela određena energijom čestice koja se nalazi na tim koordinatama. Na energije čestica dodan je gausijanski šum, što odgovara stvarnoj situaciji, u kojoj je u očitanju sa senzora uvijek prisutna nesigurnost. Tako prikazanu sliku samog *jeta*

koristimo kao ulaz u klasifikator.



Slika 5.5: Primjer prikaza *jeta*, tj. histograma energija. Osi na slici predstavljaju kutove (θ, ϕ) dok je boja svakog piksela određena energijom čestice u proizvoljnim jedinicama.

Važno je napomenuti da niti jedan realan klasifikator nije optimalan, stoga ne zadovoljava jednadžbu (5.1) u potpunosti, već je ona aproksimacija. Također, ako se naša testna funkcija jako razlikuje od prave, tada klasifikator poprima ekstremne vrijednosti (jako blizu 0 ili jako blizu 1) gdje mali pomaci, iako od velikog značaja za relaciju (5.1), nisu dovoljno kažnjeni *loss* funkcijom. Da bismo doskočili tomu, problemu pristupamo iterativno. Tada će svakom iteracijom testna funkcija biti "bliža" pravoj pa će i pomaci biti sve više kažnjeni. U svakoj iteraciji na početku učenja klasifikatora uzimamo težine s kraja učenja prethodne iteracije. Na taj način štedimo vrijeme, jer nam je potrebno manje epoha učenja te tjeramo sami klasifikator da bude što više optimalan te izbjegavamo probleme moguće pojave lokalnih minimuma.

U svakoj iteraciji uzimamo jednak broj *jetova* iz pravog skupa podataka i isti broj generiranih pomoću testne familije distribucija dobivene u prethodnoj iteraciji. Za treniranje CNN-a koristimo 75% tih podataka, dok preostalih 25% koristimo za validaciju treniranog modela.

5.4 Određivanje vjerojatnosnih distribucija

Prema izrazu (5.3) vjerojatnost da se pojavi određeni *jet* z opisan parametrima $z_i = (a_i, b_i)$, za $i \in (1, .., 7)$ za idealan klasifikator C iznosi

$$\rho(z) = \frac{C(z)}{1 - C(z)} q(z)\,, \tag{5.6}$$

pri čemu je $q(z)$ gustoća vjerojatnosti pojavljivanja mlaza z u skupu mlazova dobivenih predloženim generatorom. Uz pretpostavku da je svaki raspad međusobno nezavisan te da ovisi samo o masi čestice koja se raspada, jednadžba (5.6) može se zapisati na sljedeći način:

$$\prod_{i=1}^{7} \rho(a_i, b_i|m_i) = \frac{C(z)}{1 - C(z)} \prod_{i=1}^{7} q(a_i, b_i|m_i)\,. \tag{5.7}$$

Logaritmiranjem tog izraza slijedi:

$$\begin{aligned} &\sum_{i=1}^{7} \ln \rho(a_i, b_i|m_i) = \\ &\ln C(z) - \ln\left(1 - C(z)\right) + \sum_{i=1}^{7} \ln q(a_i, b_i|m_i)\,. \end{aligned} \tag{5.8}$$

Kako su distribucije $q(a_i, b_i|m)$ poznate, uz pretpostavku idealnog klasifikatora C, distribucije $\rho(a_i, b_i)$ mogu se odrediti rješavanjem linearnog sustava jednadžbi. No, zbog visoke dimenzionalnosti i činjenice da klasifikator nije idealan, rješavanje takvog sustava može se pokazati numerički zahtjevnim i nestabilnim. Umjesto toga, vjerojatnost $\rho(a_i, b_i|m_i)$ aproksimirana je pomoću nove neuralne mreže f na sljedeći način:

$$\rho(a, b|m) = \frac{e^{f(a,b,m)}}{\sum_{jk} e^{f(a_j, b_k, m)}}\,, \tag{5.9}$$

pri čemu indeksi j, k označavaju diskretni interval za vrijednosti parametara a i b te $f(a, b, m)$ predstavlja izlaznu vrijednost neuralne mreže f za uređenu trojku (a,b,m). Na ovaj način osigurava se da su gustoće vjerojatnosti glatke.

U ovom koraku predloženog algoritma generira se skup mlazova prema distribuciji $q(z)$ te se za svaki mlaz odredi izlazna vrijednost klasifikatora. Nakon toga slijedi proces učenja neuralne mreže f u kojem se minimizira L_2-udaljenost između lijeve i desne strane jednadžbe (5.8).

Mreža f je *feed-forward* neuralna mreža koja se sastoji od 9 potpuno povezanih

slojeva. Svi slojevi sadrže po 100 neurona osim zadnjeg koji sadrži jedan neuron. Aktivacijska funkcija slojeva je ReLU. Posljednji sloj nema aktivacijsku funkciju. Pri optimizaciji neuralne mreže f je korišten algoritam AdaM.

5.5 Opis algoritma

Na početku postavljamo testnu familiju distribucija $\rho_M(a, b)$. Konačni rezultat ne ovisi o testnoj familiji distribucija, ali što smo u početku udaljeniji od prave distribucije, trebat će nam treba više iteracija. Bitno je naglasiti da skup pravih *jetova* mora biti podskup svih *jetova* koji se mogu dobiti iz testne familije distribucija. Svaka iteracija algoritma se sastoji od nekoliko koraka koji se zatim ponavljaju:

1. Pomoću jet generatora napravimo skup $\mathcal{A}$ od 10 000 podataka (slika) koristeći trenutnu testnu distribuciju ρ_m^i.

2. Pomoću 10 000 podataka iz testne distribucije te 10 000 pravih podataka (uz napomenu da se 25% svakog skupa koristi za validaciju) treniramo klasifikator kao što je opisano u potpoglavlju 5.3.

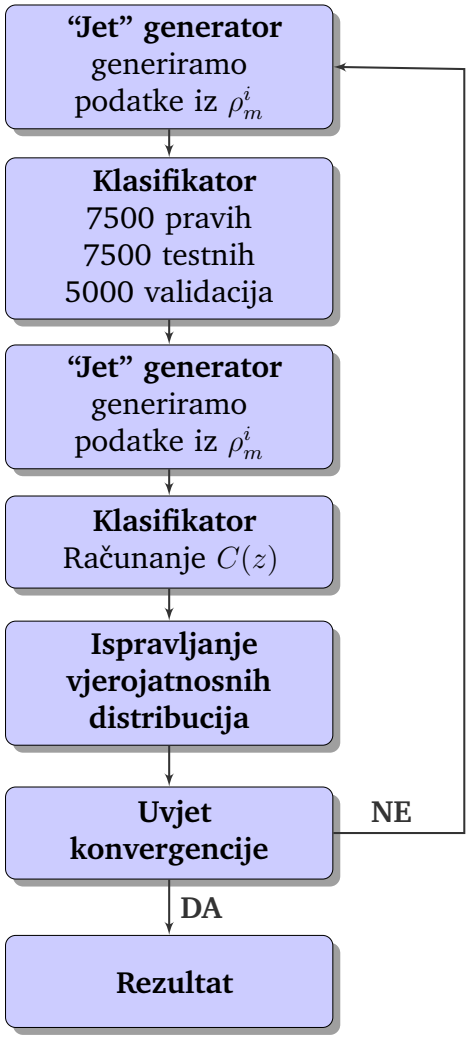
3. Pomoću *jet* generatora napravimo novi skup $\mathcal{B}$ od 10 000 podataka (slika) koristeći trenutnu testnu distribuciju.

4. Za svaki od podataka iz skupa $\mathcal{B}$ izračunamo pripadajući *score* C_{NN} pomoću klasifikatora istreniranog u 2.

5. Podatke iz 4 koristimo za ispravljanje testne distribucije na način opisan u potpoglavlju 5.4. Time dobivamo novu testnu distribuciju.

6. Provjeravamo uvjet konvergencije. Ako je zadovoljen, trenutna testna distribucija dobro opisuje pravu (tj. maksimalno koliko dozvoljava arhitektura samog klasifikatora) te je algoritam gotov. Ako uvjet nije zadovoljen, ponovimo sve prethodne korake za novu distribuciju dobivenu u 5.

Uvjet konvergencije je da klasifikator više nije u stanju raspoznati iz kojeg skupa dolaze pojedini podaci, tj. kada mu točnost kroz iteracije konvergira u 50%. Tada postižemo maksimalnu razlučivost s tom arhitekturom samog klasifikatora. U tom trenutku više se ne mijenja testna distribucija. Preciznost same metode time je ograničena rezolucijom primijenjenog klasifikatora.



Slika 5.6: Shematski prikaz algoritma korišetnog u radu.

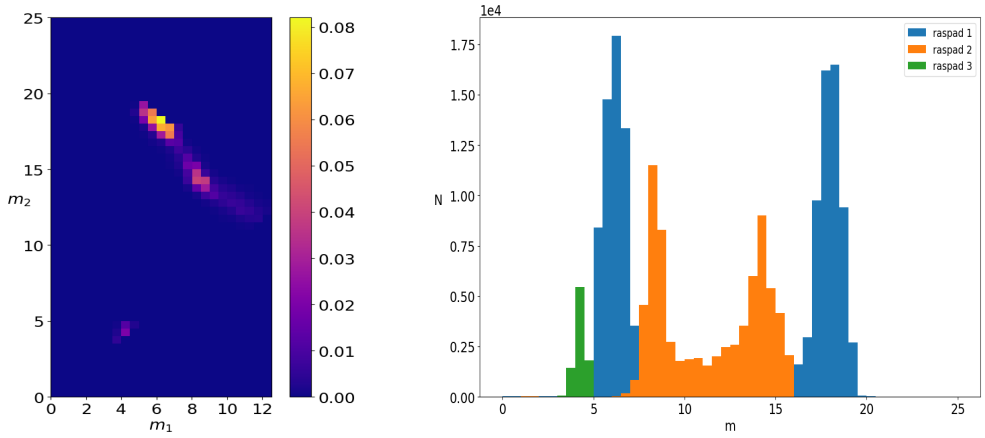
6 Rezultati

Sad primjenjujemo algoritam za rekonstrukciju distribucija raspada opisan u prethodnom poglavlju na podatke koje smo generirali po pravilima napisanim u trećem poglavlju. Prvo ćemo prezentirati rezultate kada primijenimo algoritam na skup podataka koje smo generirali po diskretnim pravilima raspada. Zatim ćemo prijeći na podatke generirane po kontinuiranim raspodjelama.

Prinuđeni rezultati su napravljeni kada treniranje samog algoritma nije još postiglo uvjet konvergencije, već je algoritam zaustavljen nakon otprilike 200 iteracija, kad točnost klasifikatora dosegne svoj minimum koji je iznad 50%. To znači da je klasifikator i dalje u stanju prepoznati razliku između skupova, međutim arhitektura mreže koja mijenja testne distribucije opisana u 5.4 sprječava daljnji napredak. Hiperparametri i arhitektura navede mreže nisu još otimizirani za ovaj problem.

6.1 Primjena na diskretne raspade

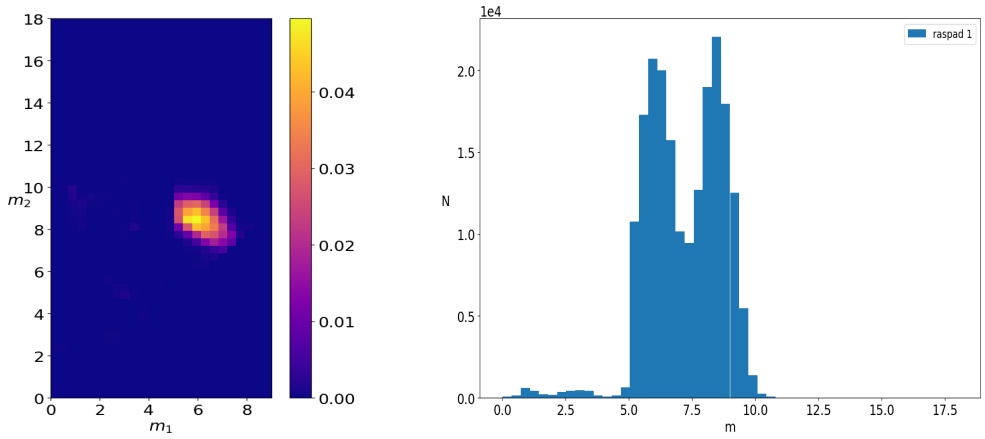
Prvo ćemo promotriti raspad čestice "majke", tj. raspad čestice mase 25. Sa slike 3.1 vidimo da se ona može raspasti na više načina: dvije čestice masa 6.1 i 18.1 s vjerojatnosti 50%, dvije čestice masa 8.4 te 14.2 s vjerojatnosti 40% ili na dvije jednake čestice masa 4.4 s vjerojatnosti 10%. Na slici 6.1 (lijevo) prikazujemo dvodimenzionalnu raspodjelu masa m_1 i m_2 dobivenu opisanim algoritmom prilikom raspada čestice mase 25.



Slika 6.1: Dvodimenzionalna distribucija dobivenih masa prilikom raspada čestice mase 25 (lijevo) i histogram masa nakon raspada čestice "majke" ($m = 25$) na 100 000 podataka (desno).

Možemo primijetiti da su se pojavile tri grupacije oko točaka $(m_1, m_2) = (4.2 \pm 0.2, 4.4 \pm 0.5)$, $(m_1, m_2) = (9 \pm 1, 14 \pm 1)$ i $(m_1, m_2) = (6.2 \pm 0.7, 17.9 \pm 0.7)$ koje vrlo dobro odgovaraju parovima vrijednosti (m_1, m_2) za sva tri moguća raspada. Na slici 6.1 (desno) vidimo raspodjelu nastalih masa raspadom čestice $m = 25$. Iz slike se jasno uočavaju vrhovi u okolini masa 4.4,6.1,8.4,14.2,18.1 na koje se čestica mase 25 po pravilima raspada. Sumirali smo vjerojatnosti za svaku od nastale tri grupacije. Grupacija oko točke $(m_1, m_2) = (4.2, 4.4)$ ima ukupnu vjerojatnost od $p = 0.05$, grupaciji oko točke $(m_1, m_2) = (9, 1)$ ukupna vjerojatnost iznosi $p = 0.38$ dok je grupaciji oko $(m_1, m_2) = (6.2, 17.9)$ ukupna vjerojatnost $p = 0.57$, što ima slično ponašanje kao i stvarni raspadi čije su vjerojatnosti redom $p = 0.1, 0.4, 0.5$. Uzmimo u obzir da u trenutku prekida algoritma, klasifikator nije još postigao uvjet konvergencije, što znači da tu razliku naš klasifikator još uvijek "vidi".

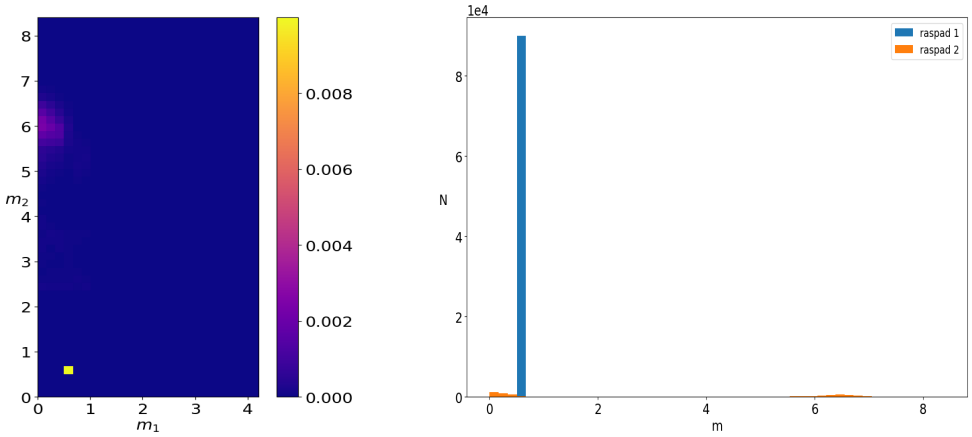
U mogućnosti smo razmotriti i daljnje raspade čestica dobivenih pomoću testne distribucije. Na slici 6.2 (lijevo) prikazana je dobivena dvodimenzionalna raspodjela masa nastalih raspadom čestice mase $m = 18.1$. Vidimo samo jednu grupaciju oko $(m_1, m_2) = (6.0 \pm 0.9, 8.5 \pm 0.8)$. To se također dobro slaže sa pravim raspadom kod kojeg se čestica $m = 18.1$ raspada na dvije čestice masa 6.1 i 18.4.



Slika 6.2: Dvodimenzionalna distribucija dobivenih masa prilikom raspada čestice mase 18.1 (lijevo) i histogram masa nakon raspada čestice ($m = 18$) na 100 000 podataka (desno).

Na slici 6.2 (desno) vidimo histogram masa nastalih raspadom čestice mase $m = 18.1$ i vidimo jasno dvije grupacije oko masa 6 i 8.5 što i očekujemo. Iz prethodne dvije analize možemo zaključiti da smo uspješno pronašli čestice sa malim vremenom života koje se neće detektirati u ovoj konstrukciji produkcije *jetova*.

Promotriti ćemo još rezultate kod raspada čestice mase $m = 8.4$. Na slici 6.2 (lijevo) prikazana je dobivena dvodimenzionalna raspodjela masa nastalih čestica. Možemo primijetiti dvije grupacije oko točaka $(m_1, m_2) = (0.59 \pm 0.01, 0.59 \pm 0.01)$ i $(m_1, m_2) = (0.1 \pm 0.2, 6.0 \pm 0.3)$. Ukupne vjerojatnosti unutar pojedinih grupacija su redom $p = 0.94$ i $p = 0.06$. Takvi rezultati približno odgovaraju pravilima opisanim na slici 3.1 u kojoj čestica mase 8 ima dva moguća raspada: prvi na dvije identične čestice mase 0.6 s vjerojatnošću 90%, a drugi na čestice mase 0.1 i 6.1. s vjerojatnošću 10%. Time zaključujemo da smo ponuđenim algoritmom u mogućnosti pronaći pravila raspada nedetektabilnih čestica. Možemo primjetiti još kako se devijacija smanjuje što je raspad više vjerojatniji.

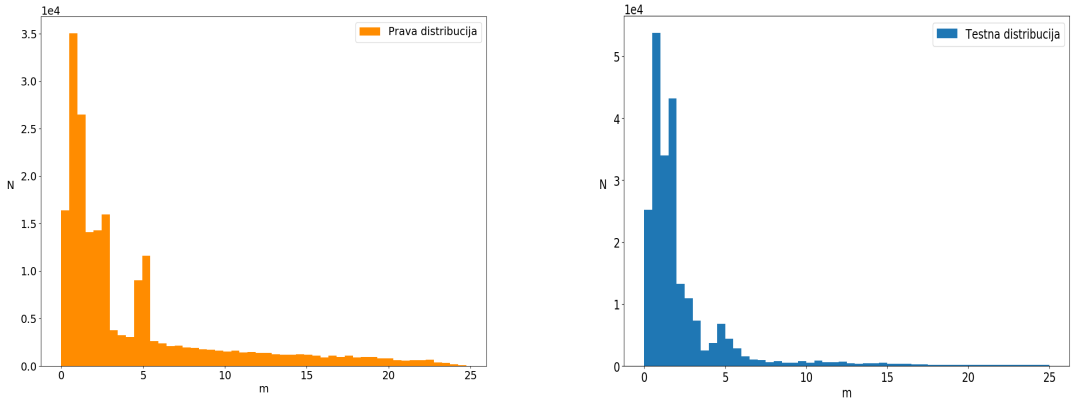


Slika 6.3: Dvodimenzionalna distribucija dobivenih masa prilikom raspada čestice mase 8.4 (lijevo), histogram masa nakon raspada čestice ($m = 8.4$) na 100 000 podataka (desno).

6.2 *Primjena na raspade s kontinuiranim distribucijama*

U ovom slučaju opet ćemo usporediti distribucije raspada za par čestica, no koristeći algoritam opisan u poglavlju 3.2. Distribucije navedene u tom poglavlju sad nazivamo pravim, dok distribucije dobivene treniranim algoritmom nazivamo testnim. Na slici 6.4 vidimo usporedbu prave distribucije (lijevo) i naše testne distribucije dobivene u trenutku kad smo zaustavili treniranje algoritma (desno).

Kod raspada čestice mase 25 do izražaja dolazi problem diskretizacije prostora parametara (a, b). Na histogramu distribucije dobivene iz "pravih" raspada jasno se

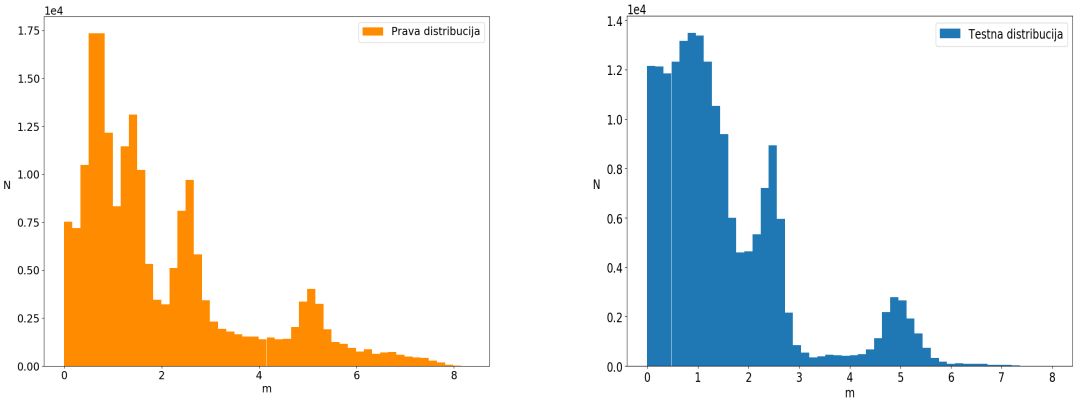


Slika 6.4: Histogram masa nakon raspada čestice (m=25) na 100 000 podataka kod prave distribucije (lijevo) i kod testne distribucije (desno).

vidi samo vrh samo zbog proizvodnje čestice mase 5, dok se vrhovi na masama ostalih stabilnih čestica ne uočavaju. Sličan problem javlja se i kod treniranja algoritma. Izgleda da se diskretizacijom parametara dodatno smanji razlučivost i s obzirom na to da su parametri a i b zapravo omjeri početne i neke od dobivenih masa, kad radimo s velikim masama razlike između parametara a i b prilikom nastanka različitih stabilnih čestica budu male. Diskretizacijom se tad distribucije u tom području dodatno zamrljaju. Također, s obzirom na to da su euklidske udaljenosti između parametara koje predstavljaju raspade na stabilne čestice male, neuronska mreža opisana u poglavlju 5.4 nema dovoljnu razlučivost da im pridoda bitno različite vrijednosti. Stoga, na slici 6.4 (desno) vidljivo da je se u dobivenoj testnoj distribuciji uočava vrh oko mase 5 te da dobivena distribucija prati pravu na većini intervala za velike mase. Međutim, zbog navedenih razloga na malim masama pojavljuju se razlike u distribucijama.

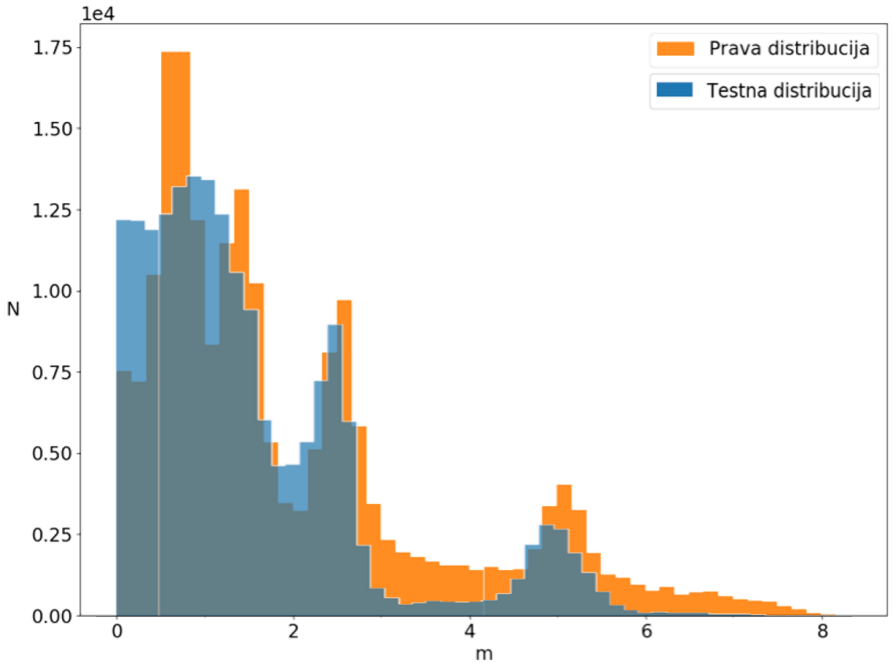
Sad ćemo pogledati raspade čestica nešto manjih masa. Na slici 6.5 vidimo distribuciju dobivenih masa za raspad čestice mase $m = 8$. Vidimo da je testna distribucija približno jednaka pravoj u okolinama masivnijih stabilnih čestica, što je zadovoljavajuće u ovom trenutku razvoja metode. Primjećujemo da se javlja i vrh u testnoj distribuciji oko stabilne čestice mase 1.4, no i dalje je razlučivost za manje masivne stabilne čestice premala.

Radi usporedbe i s obzirom na to da su obje raspodjele normirane jednako, oba histograma prikazujemo na istom grafu na slici 6.6. Jasno je vidljiva sličnost raspodjela pri masama 2.5 i 5, kao što se i očekuje iz 3.7. U testnoj distribuciji pojavljuje se



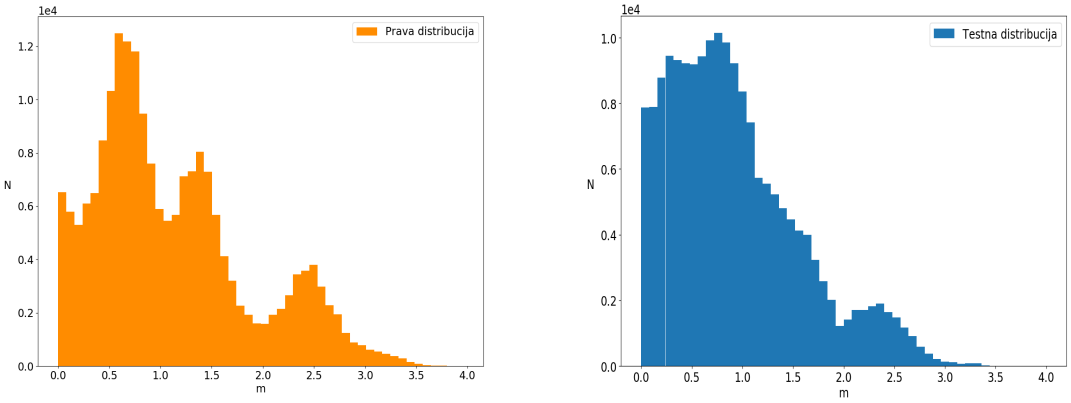
Slika 6.5: Histogram masa nakon raspada čestice (m=8) na 100 000 podataka kod prave distribucije (lijevo) i kod testne distribucije (desno).

nešto manji broj događaja oko tih masa, što je nadoknađeno većim brojem događaja na malim masama. Drugim riječima, u testnoj distribuciji kontinuirana fizikalna pozadina, koja nam ustvari nije toliko bitna za reprodukciju masa, izgleda nešto drugačije no u pravoj distribuciji. U nastavku rada u planu nam je na dobivene rezultate napraviti prilagodbu na zbroj eksponencijalno padajuće funkcije i nekog broja Gaussijana kako bismo simulirali nastanak čestica s pozadinom. Oduzimanjem pozadine, tj. eksponencijalno padajuće funkcije tad se mogu jasno izolirati čestice koje nastaju u raspadima.



Slika 6.6: Usporedba histograma masa nakon raspada čestice (m=8) na ukupno 200 000 podataka kod prave (narančaste) i testne (plave) distribucije.

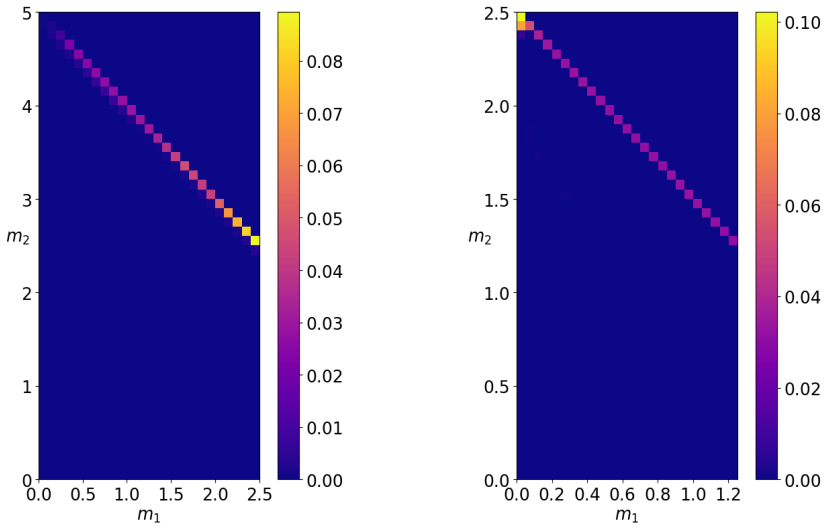
Slično ponašanje kao za česticu mase $m = 8$ uočava se i u slučaju raspada čestice mase $m = 4$, koje prikazujemo na slici 6.7. Osim izraženog vrha pri masi 2.5 u obje distribucije, pri manjim masama dolazi do smanjenja razlučivosti zbog kojeg se u testnoj distribuciji vrhovi koji su nešto jasnije izraženi u pravoj distribuciji spajaju u testnoj distribuciji. Ovakvo ponašanje je očekivano s obzirom na bliskost masa u različitim kanalima raspada i razlučivost mreže.



Slika 6.7: Histogram masa nakon raspada čestice (m=4) na 100 000 podataka kod prave distribucije (lijevo) i kod testne distribucije (desno).

Konačno, htjeli smo vidjeti kako ovaj algoritam tretira raspade stabilnih čestica. Na slici 6.8 prikazane su dvodimenzionalne raspodjele masa m_1 i m_2 dobivenih opisanim algoritmom prilikom raspada čestica masa $m = 2.5$ i $m = 5$. No, iz pravila opisanih u 3.2 znamo da se čestice masa 2.5 i 5 smatraju stabilnim, tj. da se ne raspadaju. S druge strane, na oba panela slike vidi se da postoje neki raspadi. Svi se raspadi nalaze na pravcu $m_1 + m_2 = m$, a taj pravac opisuju situaciju u kojoj se čestice nakon raspada gibaju kolinearno. S obzirom da naš zamišljeni detektor detektira samo energiju i impulse čestica možemo zaključiti da je ta situacija ekvivalentna situaciji u kojoj on detektira samo jednu česticu mase m, tj., onoj u kojoj se čestica ne raspada. U zaključku, algoritam konstrukcije se u slučaju stabilnih čestica pokazao uspješnim.

Spomenimo još jednom da je čitava analiza napravljena u trenutku kad hiperparametri algoritma nisu još do kraja optimizirani. Testne distribucije su uspješno rekonstruirale raspade na čestice velikih masa. Prikazani rezultati, s tim u vidu, daju optimističnu sliku što se tiče razvoja i uspješnosti same metode.

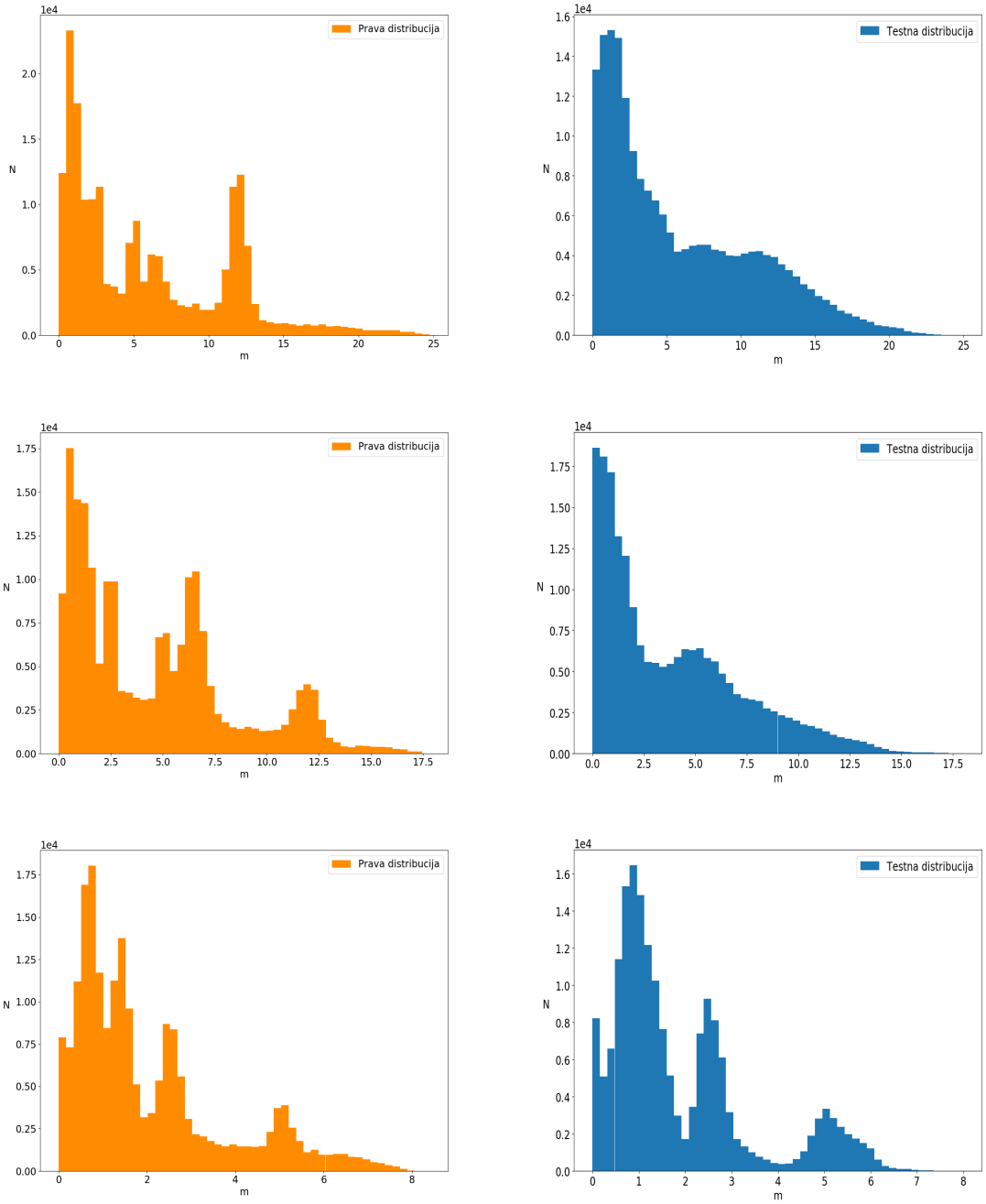


Slika 6.8: Dvodimenzionalna distribucija dobivenih masa prilikom raspada čestice mase 2.5 (lijevo) i čestice mase 5 (desno).

6.3 Primjena na raspade s kontinuiranim distribucijama i skrivenim rezonancama

Cilj u ovom potpoglavlju bio je istražiti može li navedeni algoritam rekonstrukcije razlučiti potencijalne skrivene rezonance unutar sustava koji se generira kontinuiranim raspodjelama raspada. Kao i u prethodnim slučajevima, gledati ćemo distribucije raspada za pojedine čestice. Odabrane su početne mase čestica takve da su dozvoljene obje skrivene rezonance (masa m =12.5 i 6.5). Na slici 6.9, gornja dva panela, prikazani su histogrami masa za raspade čestica mase 25 i 18 za pravu i testnu distribuciju. Za raspade čestice mase 25 u testnoj se distribuciji može slabo uočiti rezonanca na 12.5, no s velikom širinom. U raspadima čestica mase 18 na žalost se ne vide jasno skrivene rezonance, vjerojatno zbog loše razlučivosti i malih razlika u konačnom stanju, što potvrđuje ranije izveden zaključak prikazan na 3.6. Također, proizvodnja skrivene rezonance u drugom slučaju relativno je mala pa ju mreža teško prepoznaje.

Pogledali smo i kako algoritam radi kad se masa početne čestice postavi na 8. U tom slučaju jedna je skrivena rezonanca zabranjena energetski, a druga je ograničena faznim prostorom raspada jer joj je masa vrlo bliska masi početne čestice. U ovom slučaju zaključujemo da algoritam dobro rekonstruira pozadinsku distribuciju kad se zanemare skrivene rezonance.



Slika 6.9: Histogram masa nakon raspada čestice m=25 (gornji paneli), 18 (srednji paneli) i 8 (donji paneli) na 100 000 podataka kod prave distribucije (lijevo) i kod testne distribucije (desno). Obje distribucije su kontinuirane i sadrže skrivene rezonance.

7 Zaključak

U ovom radu predstavljena je metoda za nalaženje skrivenih gustoća vjerojatnosti fizikalnih sustava koji uključuju kaskadne raspade čestica. U tu svrhu razvijena su tri različita generatora događaja. Prvi od njih uključuje samo diskretne raspade određenih vjerojatnosti. Cilj metode u tom slučaju bio je odrediti moguće raspade čestica određenih masa i njihove pripadne vjerojatnosti. Na taj se način mogu pronaći čak i skrivene rezonance koje se ne pojavljuju u konačnom stanju. Dobiveni rezultati iz testne distribucije vrlo se dobro poklapaju s pravim gustoćama vjerojatnosti te su skrivene rezonantne čestice uglavnom uspješno identificirane, pogotovo u raspadima čestica veće mase. Metoda se pokazala relativno uspješnom i za drugi i treći generator, koji rade s kontinuiranim gustoćama vjerojatnosti, no u ovom slučaju pokazuje poteškoće pri identificiranju skrivenih rezonanci. Kako bi se navedeno poboljšalo, potrebno je daljnje optimiziranje neuralnih mreža.

Ovaj rad služi kao dokaz koncepta predložene metode tako da hiperparametri korištenih neuralnih mreža još uvijek nisu optimizirani, što uzrokuje nepreciznosti u konačnim rezultatima. Također, u ovom trenutku uvjet konvergencije nije još potpuno postignut zbog toga što arhitektura neuralne mreže koja opisuje distribucije nije optimizirana. Tu postoji prostor za napredak koji možemo postići pravilnim odabirom arhitekture mreže i potrebnih hiperparametara.

Prednost predložene metode je to što način na koji podaci trebaju biti reprezentirani nije određen pa se bilo kakav klasifikator koji minimizira *binary cross entropy loss* može koristiti u sklopu ove metode. To omogućava primjenu metode na realne podatke prikupljene nekim od postojećih visokoenergetskih sudarivača kao što je LHC. Iako smo u ovom radu koristili energiju čestice kao podatak iz kojeg smo generirali slike, isto tako mogli smo koristiti i transverzalni impuls čestice, koji je prirodna varijabla za opisivanje čestica uočenih u detektoru, što dodatno ukazuje na općenitost metode.

Naravno, navedena metoda u ovom trenutku ima i neke dodatne nedostatke. Npr. trenutno je ograničena razlučivost u gustoćama vjerojatnosti zbog diskretizacije prostora parametara (a, b). Povećanje razlučivosti moguće je postići finijom diskretizacijom tog prostora, no to bi uvelike otežalo numeričke proračune. Također, može se dodatno razviti i sam generator koji bi imao više stupnjeva slobode. U ovom trenutku on

posjeduje samo dva stupnja slobode koja određuje sam raspad. Kao dodatan stupanj slobode mogao bi se dodati kut raspada. Također, generator je u našem slučaju ograničen tako da se jedna čestica raspada na maksimalno dvije čestice na ljusci mase. Generator se može još dodatno poopćiti tako da omogućimo raspade i na tri čestice.

Spomenimo još da metoda razvijena u ovom radu nije ograničena na problem s kaskadnim raspadom čestica, već se može primijeniti i na druge probleme gdje je potrebno naći skrivene distribucije. Razlika bi tad bila samo u konstrukciji samog generatora. Potencijalne buduće primjene predložene metode uključuju određivanje fragmentacijskih funkcija kvarkova i gluona u mlazovima te određivanje iznosa omjera grananja za već poznate raspade. Također, ovakva metoda mogla bi pomoći pri potrazi za nepoznatim rezonancama, npr. supersimetričnim česticama.

Literatura

[1] Thomson, M. Modern Particle Physics, 1st ed. Cambride University Press, 2013.

[2] Introduction to Renormalization - QCD, University of Southampton, http://www.personal.soton.ac.uk/ab1u06/teaching/qft/qft3/lit/2007-04-20-chen.pdf, 12.4.2021.

[3] https://en.wikipedia.org/wiki/Neural_network

[4] Nielsen, M.A. Neural Networks and Deep Learning. Available online: neuralnetworksanddeeplearning.com/

[5] https://7-hiddenlayers.com/deep-learning-2/

[6] Jercic, M.; Poljak, N. Exploring the Possibility of a Recovery of Physics Process Properties from a Neural Network Model. Entropy 2020, 22, 994

[7] Kingma, D.; Ba, J. A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.