

# Razotkrivanje strukture hadrona gausijanskim procesom

Andrija Radočaj\*

mentor prof. dr. sc. Krešimir Kumerički  
Fizički odsjek, Prirodoslovno-matematički fakultet, Bijenička 32, Zagreb

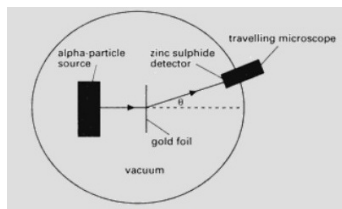
21. siječnja 2024.

## Sažetak

U ovom seminaru razmatra se jedna od metoda ekstrakcije partonskih distribucijskih funkcija određenog hadrona iz podataka o diferencijalnim udarnim presjecima više različitih procesa raspršenja nad istim hadronom. Metoda koja se proučava spada u domenu strojnog učenja i bazirana je na gausijanskim procesima. Cilj je uz pomoć jednostavnog modela i simuliranih podataka ispitati takvu metodu te saznati nešto o njenoj primjenjivosti na stvarne eksperimentalne podatke. Iznesene su teorijske osnove dubokog neelastičnog raspršenja, strojnog učenja te regresije nad gausijanskim procesima. Cijela metoda razrađena je u programskom jeziku Python uz pomoć `lsqfitgp` [1] paketa. Dobiveni rezultati obećavajući su te pokazuju da proučavani način ekstrakcije, uz poneke preinake i nadogradnje, može dobro služiti u istraživanju.

## 1. Uvod

Važnost procesa raspršenja pri razotkrivanju strukture materije prvi put je došla do izražaja 1909. godine u eksperimentu pod vodstvom Ernesta Rutherforda [2]. Naime, E. Rutherford i njegovi suradnici ispaljivali su usku zraku alfa čestica na tanku metalnu foliju i pratili kako te čestice mijenjaju svoju putanju u odnosu na početnu. Dobiveni rezultati eksperimenta pružili su dokaz da se atom sastoji od pozitivne jezgre koja sadrži većinu mase atoma.

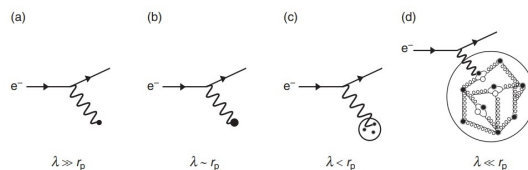


Slika 1: Shematski prikaz E. Rutherfordovog ekperimenta raspršenja [2]

---

\*aradocaj.phy@pmf.hr / andrija.radocaj124@gmail.com

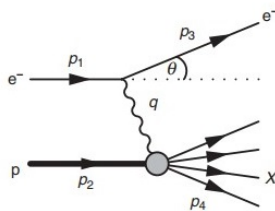
Drugim riječima, samim procesom raspršenja na atomima i poznavanjem pripadne teorijske priče saznalo se više o strukturi koja ih čini. Analogno, raspršenjem određenih čestica na protonu ili bilo kojem drugom hadronu mogu se dobiti informacije o podstrukturi koju taj hadron krije. Teorija vezana uz jedno takvo raspršenje kratko je opisana u sljedećem poglavlju. Eksperimenti raspršenja na hadronima prvi puta su se počeli izvoditi na protonima 1960-ih godina uz pomoć visokoenergetskih elektrona [2]. Na Slici 2 ispod zorno su prikazani različiti režimi navedenog raspršenja. Podstruktura mete dolazi do izražaja tek na dovoljno visokoj energiji projektila.



Slika 2: Priroda elektron-proton raspršenja ovisno o valnoj duljini virtualnog fotona [3]

## 2. Duboko neelastično raspršenje

Za razliku od elastičnog raspršenja  $e^-p \rightarrow e^-p$  koje se može opisati samo s jednom kinematičkom varijablom, kod neelastičnog raspršenja  $e^-p \rightarrow e^-X$  prikazanog na Slici 3 postoji dodatni stupanj slobode [3]. Neelastično raspršenje stoga se opisuje s dvije nezavisne kinematičke varijable.



Slika 3: Neelastično raspršenje elektrona na protonu [3]

Neke od kinematičkih varijabli koje se koriste su Bjorkenov  $x$  te varijable  $Q^2$ ,  $y$  i  $\nu$ :

$$Q^2 = -q^2 = -(p_1 - p_3)^2, \quad (1)$$

$$x = \frac{Q^2}{2p_2 \cdot q} \quad (0 \leq x \leq 1), \quad (2)$$

$$y = \frac{p_2 \cdot q}{p_2 \cdot p_1} \quad (0 \leq y \leq 1), \quad (3)$$

$$\nu = \frac{p_2 \cdot q}{m_p}, \quad (4)$$

gdje su  $p_1$ ,  $p_2$ ,  $p_3$  te  $q$  četveroimpulsi označeni na Slici 3, a  $m_p$  je masa protona.

Duboko neelastično raspršenje elektrona na određenom hadronu čine elastična raspršenja elektrona na kvarkovima. Za jedno takvo elastično raspršenje na kvarku može se izvesti izraz za diferencijalni udarni presjek koji ima oblik:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2 Q_q^2}{Q^4} \left[ (1-y) + \frac{y^2}{2} \right], \quad (5)$$

gdje je  $Q_q$  naboj kvarka na kojem se odvija raspršenje.

Partonska distribucijska funkcija  $q^p(x)$  za kvark određenog okusa  $q$  definira se tako da  $q^p(x)\delta x$  predstavlja broj kvarkova okusa  $q$  unutar hadrona kojima je udio impulsa hadrona između  $x$  i  $x + \delta x$  [3]. Izraz za elastično raspršenje elektrona na svim kvarkovima određenog okusa kojima je udio impulsa hadrona u spomenutim granicama ima stoga oblik:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[ (1-y) + \frac{y^2}{2} \right] \times Q_i^2 q_i^p(x) \delta x. \quad (6)$$

Izraz dobiven iznad potom se podijeli sa  $\delta x$  te se zbroje svi okusi kvarkova koji postoje u određenom hadronu pa slijedi:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[ (1-y) + \frac{y^2}{2} \right] \sum_i Q_i^2 q_i^p(x). \quad (7)$$

Partonske distribucijske funkcije govore o raspodjeli impulsa među kvarkovima unutar hadrona te su stoga važan izvor informacije o samom hadronu. U izrazu (7) iznad eksperimentalno mjerljiva veličina je diferencijalni udarni presjek te je uz pomoć dobivenih rezultata mjerenja cilj ekstrahirati partonske distribucijske funkcije. Postoje različite metode kojima se tom problemu može pristupiti, no ovaj seminar bavi se metodom strojnog učenja baziranoj na gausijanskim procesima.

### 3. Strojno učenje

Strojno učenje definira se kao način programiranja računala tako da ono uz pomoć prethodnog iskustva ili podatkovnih primjera optimizira neki kriterij uspješnosti [4]. Cilj je moći dobro opisati podatke koji računalu trenutno nisu dostupni uz pomoć podataka koji mu se nude.

Kod strojnog učenja raspolaze se modelom koji se definira do na neke parametre. Spomenuti parametri modela nastoje se optimizirati tako da model što bolje opisuje dostupne podatke. Ipak, uzimajući u obzir cilj kojemu se teži, model ne smije biti previše složen te opisivati dostupne podatke savršeno bez mogućnosti dobre generalizacije na podatke koji mu nisu dostupni.

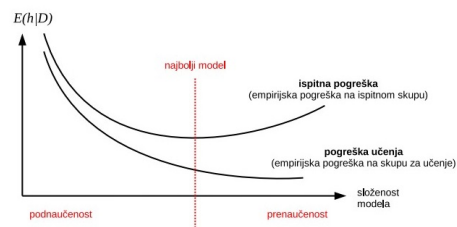
#### 3.1 Osnovna podjela strojnog učenja

Budući da učenje podrazumijeva interakciju između učenika i okoline, zadaci učenja se mogu podijeliti prema prirodi te interakcije [5]. Postoje dvije kategorije strojnog učenja: učenje pod nadzorom i učenje bez nadzora. Kao ilustrativni primjer razlike između spomenute dvije kategorije može se uzeti zadatak učenja otkrivanja nepoželjnih mailova. U slučaju učenja pod nadzorom učenik dobiva skup mailova za vježbu gdje je svaki mail označen s poželjan ili nepoželjan. Na

temelju takve vježbe učenik treba naučiti pravilo za označivanje novih pristizućih mailova. Drugim riječima, učenje pod nadzorom je proces korištenja iskustva pri razvijanju stručnosti. Osim problema klasifikacije, kao što je sortiranje mailova, u ovu kategoriju spadaju još problemi regresije. Metoda strojnog učenja kojom se ovaj seminar služi pri ekstrakciji partonskih distribucijskih funkcija iz podataka diferencijalnog udarnog presjeka u sebi sadrži upravo regresijski problem. Za razliku od učenja pod nadzorom, učenje bez nadzora je oblik učenja gdje sve što učenik dobije je veliki broj mailova bez oznaka i njegov je zadatak otkriti one nepoželjne [5]. Dakle, učenik u ovom slučaju obrađuje dobivene podatke s ciljem da se dosjeti nekog sažetka tih podataka.

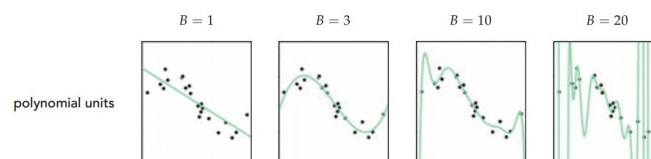
### 3.2 Optimizacija modela

Pri optimizaciji parametara odabranog modela potrebno je poznavati neku funkciju gubitka koju će program nastojati minimizirati te algoritam po kojemu će to činiti. Funkcija gubitka može se izvršiti na skupu za učenje, ali i na ispitnom skupu koji mu pri učenju nije dostupan.



Slika 4: Teorijska ovisnost pogreške o složenosti modela na skupu za učenje i na ispitnom skupu [4]

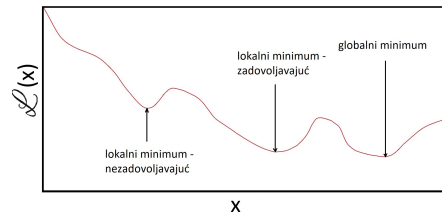
Cilj modela je da dobro opisuje neviđene podatke pa su, prema Slici 4 iznad, optimalni parametri modela oni kojima model minimizira funkciju gubitka na ispitnim primjerima, a ne na primjerima za učenje [4]. Ipak, budući da u procesu učenja algoritam nema pristup ispitnom skupu, jedino što on može učiniti je minimizirati funkciju gubitka na skupu za učenje. Iz navedenog razloga pri učenju može doći do nepoželjne prenaučnosti. Zoran prikaz režima podnaučenosti i prenaučnosti nalazi se na Slici 5 ispod.



Slika 5: Prilagodba polinoma različitih stupnjeva na mjerene podatke [6]

Problemu prenaučnosti može se po potrebi doskočiti takozvanom regularizacijom koja transformira funkciju gubitka na način: funkcija gubitka  $\leftarrow$  funkcija gubitka + složenost modela. Transformiranoj funkciji gubitka cilj je tada algoritmom minimizacije pronaći parametre koji odgovaraju njenom globalnom minimumu. Naime, tako optimizirani parametri modela trebali bi dobro opisivati skup za učenje pritom ne čineći model presloženim i time ugrožavajući generalizaciju na ispitne podatke.

Ipak, ponekad je regularizaciju moguće izbjeći bez posljedica. Naime, potrebno je samo uzeti u obzir da algoritmi minimizacije kreću s nekom zadanom parametarskom točkom, rade u koracima te pri zaustavljanju ne razlikuju lokalni minimum od globalnog. Zoran prikaz takvog rada jednog jednostavnog algoritma minimizacije nalazi se na Slici 8 ispod. Nadalje, ako je funkcija gubitka jednoparametarska njena ovisnost o parametru se može nacrtati s ciljem uočavanja položaja minimuma. Pomnim odabirom početne parametarske točke jednostavna verzija algoritma minimizacije može se zaustaviti na obližnjem lokalnom minimumu, a ne na globalnom. Prikladan lokalni minimum je onaj koji ne čini model presloženim, a funkcijska vrijednost mu je bliska funkcijskoj vrijednosti globalnog minimuma. Opisanim pristupom moguće je bez regularizacije izbjeći prenaučenosť koju bi potencijalno izazvao globalni minimum funkcije gubitka. Ilustrativni prikaz odabira zadovoljavajućeg lokalnog minimuma prikazan je na Slici 6 ispod.



Slika 6: Ilustrativni prikaz ovisnosti neke funkcije gubitka o jednom parametru

S druge strane, funkciju gubitka koja ovisi o više parametara nije moguće vizualizirati. Ipak, slična stvar se može postići višestrukim pokušajima odabira početne parametarske točke. Naime, pogledom na prilagodbu podataka nakon svakog završetka algoritma minimizacije može se procijeniti jeli dobivena prilagodba zadovoljavajuća. U ovom seminaru upravo koristeći taj pristup neće biti potrebe za regularizacijom koja u Python paketu `lsqfitgp` do trenutka pisanja ovog seminarskog rada nije dostupna opcija.

Postoji više algoritama koji se koriste za minimizaciju neke funkcije gubitka. Neki od njih su gradijentni spust, stohastički gradijentni spust te njihove različite varijacije i nadogradnje [6]. Spomenute varijacije i nadogradnje ova dva pristupa služe kako bi se izbjeglo zapinjanje u lokalnim minimumima te smanjilo vrijeme koje treba algoritmu da postigne globalni minimum. Ovaj seminar služi se `lsqfitgp` paketom Pythona koji u najnovijoj verziji kao algoritam minimizacije sadrži samo obični gradijentni spust bez regularizacije funkcije gubitka. Navedeno će se pokazati dovoljnim za ilustraciju onoga što se želi postići.

### 3.3 Gradijentni spust

Gradijentni spust je osnovni algoritam za minimizaciju funkcije koja je definirana do na parametar. Njegova shema opisana je ispod i bazira se na činjenici da funkcija najbrže opada u smjeru obrnutom od gradijenta:

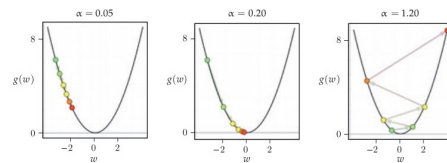
**Ulaz:** skup za učenje  $(X, Y)$ , diferencijabilna funkcija gubitka  $\mathcal{L}(X, Y, \mathbf{w})$ , stopa učenja  $\eta^{(i)}$ , početna točka parametra  $\mathbf{w}^{(0)} = (w_1^{(0)}, \dots, w_n^{(0)})$

1.  $i=0$
2. **while** uvjet zaustavljanja nije ispunjen **do**
3.     Računanje  $\nabla_{\mathbf{w}}\mathcal{L}(X, Y, \mathbf{w}^{(i)})$
4.     Nadogradnja  $\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} - \eta^{(i)}\nabla_{\mathbf{w}}\mathcal{L}(X, Y, \mathbf{w}^{(i)})$
5.      $i \leftarrow i + 1$
6. **end while.**

Mogući uvjeti zaustavljanja while petlje su:

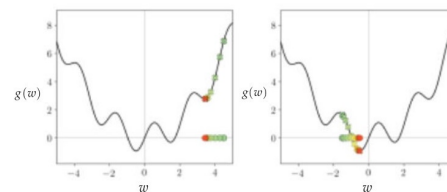
- $\|\nabla_{\mathbf{w}}\mathcal{L}(X, Y, \mathbf{w}^{(i)})\|_2 \approx 0 \rightarrow$  mala norma gradijenta,
- $\|\mathbf{w}^{(i+1)} - \mathbf{w}^{(i)}\|_2 \approx 0 \rightarrow$  mala promjena parametara od jedne do druge epohe,
- $|\mathcal{L}(X, Y, \mathbf{w}^{(i+1)}) - \mathcal{L}(X, Y, \mathbf{w}^{(i)})| \approx 0 \rightarrow$  mala promjena funkcije gubitka od jedne do druge epohe.

Stopa učenja navedena u algoritmu mora biti pomno izabrana. Naime, premala stopa učenja dovela bi do spore konvergencije, a prevelika stopa učenja mogla bi uzrokovati da algoritam uopće ne konvergira. Opisana problematika prikazana je na Slici 7 ispod.



Slika 7: Ilustrativni prikaz ovisnosti rada algoritma gradijentnog spusta o zadanoj stopi učenja od lijeva na desno: premala stopa, optimalna stopa, prevelika stopa [6]

Navedeni uvjeti zaustavljanja while petlje i Slika 8 ispod ukazuju da se, ovisno o izboru početne parametarske točke, minimizacijom nekonveksne funkcije algoritam može zaustaviti ili na nekom od lokalnih minimuma ili na onom globalnom.



Slika 8: Ilustrativni prikaz rada algoritma gradijentnog spusta po epohama za dvije različite početne parametarske točke [6]

## 4. Gausijanski procesi

Gausijanski proces je skup nasumičnih varijabli od kojih svaki njihov konačan broj čini multivarijantnu Gaussovu raspodjelu, poopćenje Gaussove raspodjele na više dimenzija [7]. Multivarijantna Gaussova raspodjela za D-dimenzionalni nasumični vektor

$\mathbf{f}=(f_1,\dots,f_D)^T=(f(\mathbf{x}_1),\dots,f(\mathbf{x}_D))^T$  označuje se kao  $\mathbf{f}\sim\mathcal{N}_D(\boldsymbol{\mu},\boldsymbol{\Sigma})$ , gdje je

$\boldsymbol{\mu}=\mathbf{E}[\mathbf{f}]=(\mathbf{E}(f_1),\dots,\mathbf{E}(f_D))^T=(\mathbf{E}[f(\mathbf{x}_1)],\dots,\mathbf{E}[f(\mathbf{x}_D)])^T\equiv(\mu(\mathbf{x}_1),\dots,\mu(\mathbf{x}_D))^T$  srednja vrijednost i

$\Sigma_{i,j}=\mathbf{E}[(f(\mathbf{x}_i)-\mu_i)(f(\mathbf{x}_j)-\mu_j)]=\text{Cov}[f(\mathbf{x}_i),f(\mathbf{x}_j)]=k(\mathbf{x}_i,\mathbf{x}_j)$  funkcija kovarijance.

Uz prethodno uvedene oznake gustoća multivarijantne Gaussove raspodjele ima sljedeći oblik [8]:

$$\mathcal{N}_D(\mathbf{f}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{f}-\boldsymbol{\mu})\right]. \quad (8)$$

Gausijanski proces je dakle u potpunosti određen svojom srednjom vrijednošću i funkcijom kovarijance [7]. Definira se srednja vrijednost funkcije  $\mu(\mathbf{x})$  i funkcija kovarijance  $k(\mathbf{x},\mathbf{x}')$  stvarnog procesa  $f(\mathbf{x})$  kao što je navedeno u tekstu iznad:

$$\begin{aligned} \mu(\mathbf{x}) &= \mathbf{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbf{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))] \end{aligned} \quad (9)$$

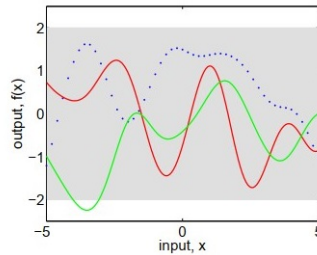
te se tada Gausijanski proces piše kao:

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (10)$$

Funkcija kovarijance koja se često koristi kod Gausijanskih procesa je takozvana kvadrirana eksponencijalna funkcija:

$$k(\mathbf{x}, \mathbf{x}') = \text{exp. quad}(\mathbf{x}, \mathbf{x}'; l) = \exp\left(-\frac{1}{2l^2}|\mathbf{x} - \mathbf{x}'|^2\right). \quad (11)$$

Na primjer, generiranjem nasumičnog Gaussovog vektora koristeći prethodno spomenutu funkciju kovarijance sa skalom  $l=1$  te crtanjem generiranih vrijednosti kao funkcije ulaznih vrijednosti dobije se Slika 9 ispod.



Slika 9: Tri primjera nasumičnog Gaussovog vektora generiranog iz raspodjele  $\mathbf{f}_* \sim \mathcal{N}(\mathbf{0}, \text{exp. quad}(X_*, X_*; l=1))$  [7]

## 4.1 Regresija nad gausijanskim procesima

Dakako, crtanje nasumičnih funkcija koje ne predstavljaju neke podatke nije od velike koristi. Cilj je dakle uključiti znanje koje skup za učenje donosi o funkciji [7]. Dostupni podatci su  $\{(\mathbf{x}_i, y_i) | i=1, \dots, n\}$ , gdje je  $y_i = f(\mathbf{x}_i) + \varepsilon$ . Kada se stvarni proces modelira kao gausijanski proces srednja vrijednost mu se pretpostavlja  $\boldsymbol{\mu} = \mathbf{0}$  te se odabire neka funkcija kovarijance  $k(\mathbf{x}, \mathbf{x}')$  koja ga treba opisati. Pretpostavkom o aditivnom nezavisno raspodijeljenom gausijanskom šumu  $\varepsilon$  sa varijancom  $\sigma_n^2$  slijede izrazi ispod:

$$\text{cov}(y_p, y_q) = k(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq} \quad \text{ili} \quad \text{cov}(\mathbf{y}) = K(X, X) + \sigma_n^2 I. \quad (12)$$

Multivarijantna Gaussova raspodjela srednjih vrijednosti mjerenja  $\mathbf{y}$  i funkcijskih vrijednosti na ispitnim točkama  $\mathbf{f}_*$  tada ima oblik:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \quad (13)$$

gdje su zvjezdicom označeni elementi skupa čije je ponašanje cilj odrediti.

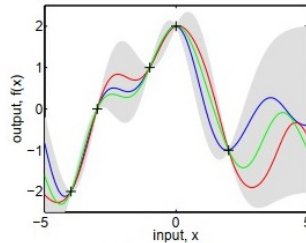
Izvodom uvjetne raspodjele iz jednadžbe iznad dolazi se do ključnih jednadžbi predviđanja za regresiju gausijanskih procesa:

$$\mathbf{f}_* | X, \mathbf{y}, X_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)), \quad \text{gdje je} \quad (14)$$

$$\bar{\mathbf{f}}_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y} \quad \text{te} \quad (15)$$

$$\text{cov}(\mathbf{f}_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*). \quad (16)$$

Naravno, postupak za mjerenja  $\mathbf{y}$  bez šuma dobije se postavljanjem  $\sigma_n^2 = 0$  te uz takvu vrstu mjerenja i funkciju kovarijance  $\text{exp. quad}(\mathbf{x}, \mathbf{x}'; l=1)$  slijede primjeri funkcija sa Slike 10 ispod.



Slika 10: Tri primjera nasumičnog Gaussovog vektora generiranog uz pomoć izraza (14)-(16) za mjerenja  $\mathbf{y}$  bez šuma i  $k(\mathbf{x}, \mathbf{x}') = \text{exp. quad}(\mathbf{x}, \mathbf{x}'; l=1)$  [7]

Pri regresiji je cilj što bolje opisati mjerene podatke biranjem najboljih kandidata za vrijednosti parametara izabrane funkcije kovarijance. Ipak, kao što je napomenuto u prethodnom poglavlju, raznim metodama pri učenju treba pripaziti da nebi došlo do prenaučnosti. U slučaju funkcije kovarijance  $\text{exp. quad}(\mathbf{x}, \mathbf{x}'; l)$  model je složeniji što je skala  $l$  manja jer skala  $l$  nosi informaciju o tome koliko moraju biti odmaknute dvije vrijednosti funkcije da bi one postale nekorelirane [7]. Prilagodba parametara funkcije kovarijance na mjerene podatke postiže se gradijentnim spustom na funkciji gubitka specifičnoj za regresiju nad gausijanskim procesom.



## 4.2 Funkcija gubitka i njezin gradijent

Proces od interesa, kao što je već spomenuto, modelira se kao gausijanski proces sa srednjom vrijednošću  $\boldsymbol{\mu}=\mathbf{0}$  i izabranom funkcijom kovarijance  $k(\mathbf{x},\mathbf{x}';\mathbf{w})$  parametra  $\mathbf{w}$ . Cilj je izračunati graničnu vjerojatnost da takav model opisuje skup za učenje. Pri računanju te vjerojatnosti treba uzeti u obzir sve funkcijske vektore s  $\mathbf{x}_i$  koordinatama skupa za učenje koji se mogu izgenerirati pretpostavljenim modelom,  $\mathbf{f}|X\sim\mathcal{N}(\mathbf{0}_n,K)$  [7]. Umnoškom vjerojatnosti za ostvarenje pojedinog funkcijskog vektora i vjerojatnosti da on opisuje proces od interesa te najavljenom sumacijom po svim takvim vektorima modela, za graničnu vjerojatnost slijedi izraz:

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|\mathbf{f},X)p(\mathbf{f}|X)d\mathbf{f}. \quad (17)$$

Izraz za prirodni logaritam granične vjerojatnosti da zadani model opisuje skup za učenje tada je uz  $p(\mathbf{y}|\mathbf{f},X)=\mathcal{N}(\mathbf{y}|\mathbf{f},\sigma_n^2\mathbf{I})$  jednak:

$$\ln[p(\mathbf{y}|X,\mathbf{w})] = -\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2\mathbf{I})^{-1}\mathbf{y} - \frac{1}{2}\ln|K + \sigma_n^2\mathbf{I}| - \frac{n}{2}\ln(2\pi). \quad (18)$$

Funkciju u izrazu (18) iznad cilj je optimizacijom parametra  $\mathbf{w}$  maksimizirati što bi značilo da se za funkciju gubitka može uzeti  $\mathcal{L}(\mathbf{w})=(-1)\cdot\ln[p(\mathbf{y}|X,\mathbf{w})]$ .

Kako bi se mogao provesti algoritam gradijentnog spusta potrebno je poznavati gradijent funkcije gubitka  $\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w})=(-1)\cdot\nabla_{\mathbf{w}}\ln[p(\mathbf{y}|X,\mathbf{w})]$  koji se računa uz pomoć izraza [7]:

$$\frac{\partial}{\partial w_j}\ln[p(\mathbf{y}|X,\mathbf{w})] = \frac{1}{2}\mathbf{y}^T K^{-1} \frac{\partial K}{\partial w_j} K^{-1} \mathbf{y} - \frac{1}{2}tr\left(K^{-1} \frac{\partial K}{\partial w_j}\right). \quad (19)$$

## 5. Ekstrakcija partonskih distribucijskih funkcija

Prvi korak u ekstrakciji partonskih distribucijskih funkcija iz podataka diferencijalnog udarnog presjeka je modeliranje tih funkcija kao nezavisnih gausijanskih procesa. Budući da se linearnom transformacijom gausijanskih procesa dobiva novi takav proces, povratkom na izraz (7) slijedi da se takvim modeliranjem partonskih distribucijskih funkcija tada istovremeno na isti način modelira i diferencijalni udarni presjek. Opisanom metodom regresije gausijanskih procesa moguće je stoga napraviti prilagodbu nad mjerenim podacima diferencijalnog udarnog presjeka za neki proces raspršenja na određenom hadronu. Nadalje, ukoliko određeni hadron sadrži  $n$  različitih kvarkova i  $m$  različitih gluona tada je potrebno poznavati podatke diferencijalnog udarnog presjeka za  $(n+m)$  različitih procesa raspršenja nad istim hadronom da bi ekstrakcija svih partonskih distribucijskih funkcija bila moguća. Naime, nakon prilagodbe na podatke diferencijalnih udarnih presjeka dobije se sustav od  $(n+m)$  jednadžbi s  $(n+m)$  nepoznanica koji ima jedinstveno rješenje.

### 5.1 Pojednostavljenje problema

Za ilustraciju učinkovitosti odabrane metode ekstrakcije mogu se uvesti određena pojednostavljenja koja predstavljaju malo drugačiju fiziku od one koja u prirodi vrijedi. Prvo pojednostavljenje je zadržavanje na samo dvije partonske distribucijske funkcije o kojima ovisi diferencijalni udarni presjek. Za ekstrakciju tih funkcija potrebno je stoga poznavati podatke o diferencijalnom udarnom presjeku za samo dva procesa raspršenja. Drugo pojednostavljenje je smanjenje broja kinematičkih

varijabli na jednu. Dva procesa raspršenja koja će se proučavati opisana su dakle pojednostavljenim izrazima za diferencijalni udarni presjek. Za potrebe ovog seminarskog rada njihove ovisnosti o partonskim distribucijskim funkcijama zadane su izrazima ispod:

$$\frac{d\sigma_1}{dx} = h_1(x)[0] \cdot \text{dquark}_{\text{truth}}(x) + h_1(x)[1] \cdot \text{gluon}_{\text{truth}}(x), \quad (20)$$

$$\frac{d\sigma_2}{dx} = 1 + h_2(x)[0] \cdot \text{dquark}_{\text{truth}}(x) + h_2(x)[1] \cdot \text{gluon}_{\text{truth}}(x), \quad (21)$$

gdje  $h_1(x)$  i  $h_2(x)$  predstavljaju poznate takozvane amplitude tvrdog raspršenja, dok  $\text{dquark}_{\text{truth}}(x)$  i  $\text{gluon}_{\text{truth}}(x)$  partonske distribucijske funkcije koje je cilj ekstrahirati. U opisanom pojednostavljenju ovog seminarskog rada spomenute amplitude zadane su na sljedeći način:

$$h_1(x) = \left[ \frac{1}{1+x}, \frac{x}{1+x} \right], \quad (22)$$

$$h_2(x) = \left[ \frac{x}{1+x^2}, \frac{1}{1+x^2} \right]. \quad (23)$$

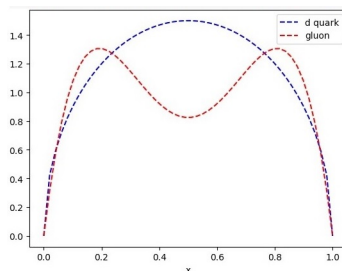
## 5.2 Simulacija eksperimentalnih podataka

S ciljem simuliranja eksperimentalnih podataka diferencijalnih udarnih presjeka zadaje se određena ovisnost partonskih distribucijskih funkcija o kinematičkoj varijabli  $x$  kao da je poznata. U ovom seminarskom radu partonske distribucijske funkcije zadane su sljedećim izrazima:

$$\text{dquark}_{\text{truth}}(x) = 3\sqrt{x \cdot (1-x)}, \quad (24)$$

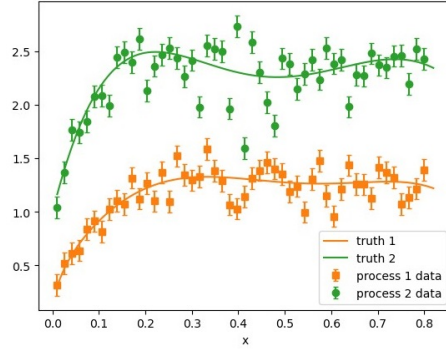
$$\text{gluon}_{\text{truth}}(x) = 6x \cdot (1-x) \cdot (1 + 0.3 \cdot C_2^{(1.5)}(2x-1)), \quad (25)$$

gdje je  $C_n^{(\alpha)}$  Gegenbauerov polinom [9].



Slika 11: Ovisnosti partonskih distribucijskih funkcija zadane izrazima (24)-(25)

Nakon zadavanja partonskih distribucijskih funkcija slijedi računanje diferencijalnih udarnih presjeka prema izrazima (20)-(23) te se dobivenim rezultatima dodaje šum. Uz pomoć tako simuliranih podataka sa Slike 12 cilj je natrag dobiti što bolju replikaciju početno zadanih partonskih distribucijskih funkcija sa Slike 11.



Slika 12: Simulirani podatci diferencijalnog udarnog presjeka za zadana dva procesa raspršenja

Nakon usporedbe ekstrahiranih partonskih distribucijskih funkcija sa onim početno zadanim bit će moguće nešto zaključiti o primjenjivosti proučavane metode ekstrakcije na stvarne eksperimentalne podatke.

### 5.3 Matrica kovarijance

Pri izračunu funkcije gubitka specifične za regresiju gausijanskih procesa zahtijeva se poznavanje srednjih vrijednosti mjerenih podataka  $\mathbf{y}$  i šuma  $\sigma_n^2$  te matrice kovarijance  $\mathbf{K}(\mathbf{X}, \mathbf{X}; \mathbf{w})$  kojom je cilj opisati zadani proces. U pojednostavljenom problemu ovog seminara diferencijalni udarni presjeci dvaju procesa raspršenja su linearne transformacije dvaju partonskih distribucijskih funkcija. Pri modeliranju tih funkcija kao nezavisnih gausijanskih procesa srednje vrijednosti  $\boldsymbol{\mu}=0$  potrebno je svakom pojedinom odabrati funkciju kovarijance koja ga predstavlja. Diferencijalni udarni presjek za svaki od procesa raspršenja tada isto postaje gausijanski proces sa funkcijom kovarijance koja ovisi o funkcijama kovarijance modeliranih partonskih distribucijskih funkcija. Naime, ako je odabrana funkcija kovarijance jedne modelirane partonske distribucijske funkcije  $\text{exp.quad}(\mathbf{x}, \mathbf{x}'; l_1)$  i druge  $\text{exp.quad}(\mathbf{x}, \mathbf{x}'; l_2)$  slijedi da je funkcija kovarijance za prvi proces raspršenja opisan izrazom (20):

$$\text{cov}_1(x, x') = h_1(x)[0] \cdot h_1(x')[0] \cdot \text{exp.quad}(x, x'; l_1) + h_1(x)[1] \cdot h_1(x')[1] \cdot \text{exp.quad}(x, x'; l_2). \quad (26)$$

S druge strane, za drugi proces raspršenja opisan izrazom (21) potrebno je simulirane podatke pomaknuti za -1 kako bi onda i taj proces mogao biti modeliran kao gausijanski kojemu se srednja vrijednost pretpostavlja  $\boldsymbol{\mu}=0$ . Funkcija kovarijance drugog procesa raspršenja ima oblik:

$$\text{cov}_2(x, x') = h_2(x)[0] \cdot h_2(x')[0] \cdot \text{exp.quad}(x, x'; l_1) + h_2(x)[1] \cdot h_2(x')[1] \cdot \text{exp.quad}(x, x'; l_2). \quad (27)$$

S obzirom na to da ovise o istim partonskim distribucijskim funkcijama, procesi raspršenja od interesa međusobno su zavisni te stoga postoji i funkcija kovarijance koja opisuje njihov odnos:

$$\begin{aligned} cov_{12}(x, x') &= h_1(x)[0] \cdot h_2(x')[0] \cdot \mathbf{exp} \cdot \mathbf{quad}(x, x'; l_1) \\ &\quad + h_1(x)[1] \cdot h_2(x')[1] \cdot \mathbf{exp} \cdot \mathbf{quad}(x, x'; l_2), \end{aligned} \quad (28)$$

$$\begin{aligned} cov_{21}(x, x') &= h_2(x)[0] \cdot h_1(x')[0] \cdot \mathbf{exp} \cdot \mathbf{quad}(x, x'; l_1) \\ &\quad + h_2(x)[1] \cdot h_1(x')[1] \cdot \mathbf{exp} \cdot \mathbf{quad}(x, x'; l_2). \end{aligned} \quad (29)$$

Srednje vrijednosti mjerenih podataka prvog procesa  $\mathbf{y}_1$  i drugog procesa  $\mathbf{y}_2$  mogu se zajedno staviti u matricu  $\mathbf{y}$  koja predstavlja srednje vrijednosti svih mjerenih podataka:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1(x_1) \\ \dots \\ \mathbf{y}_1(x_n) \\ \mathbf{y}_2(x_1) \\ \dots \\ \mathbf{y}_2(x_n) \end{bmatrix}. \quad (30)$$

Pripadna matrica funkcije kovarijance tada ima sljedeći oblik:

$$K = \begin{bmatrix} cov_1(x_1, x_1) & \dots & cov_1(x_1, x_n) & cov_{12}(x_1, x_1) & \dots & cov_{12}(x_1, x_n) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ cov_1(x_n, x_1) & \dots & cov_1(x_n, x_n) & cov_{12}(x_n, x_1) & \dots & cov_{12}(x_n, x_n) \\ cov_{21}(x_1, x_1) & \dots & cov_{21}(x_1, x_n) & cov_2(x_1, x_1) & \dots & cov_2(x_1, x_n) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ cov_{21}(x_n, x_1) & \dots & cov_{21}(x_n, x_n) & cov_2(x_n, x_1) & \dots & cov_2(x_n, x_n) \end{bmatrix} \quad (31)$$

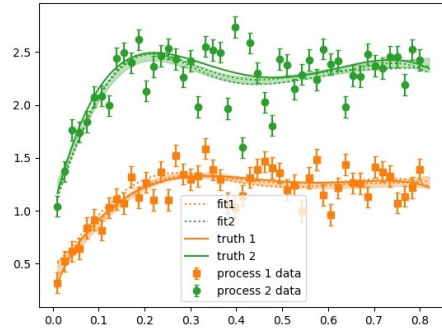
te ovisi o parametrima  $l_1$  i  $l_2$ , odnosno o vektoru  $\mathbf{w}$  koji se postavlja na način:

$$\mathbf{w} = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}. \quad (32)$$

U ovom trenutku moguće je uz pomoć danih  $\mathbf{y}$  i  $K$  izračunati prirodni logaritam granične vjerojatnosti iskazan u izrazu (18) čija negativna vrijednost predstavlja funkciju gubitka. Optimizacijom slobodnih parametara matrice  $K$  ( $l_1$  i  $l_2$ ) gradijentnim spustom na funkciji gubitka, pritom pazeći na sve što je spomenuto za problem prenaučivosti, dobije se najbolji opis simuliranih podataka za dva procesa raspršenja od interesa.

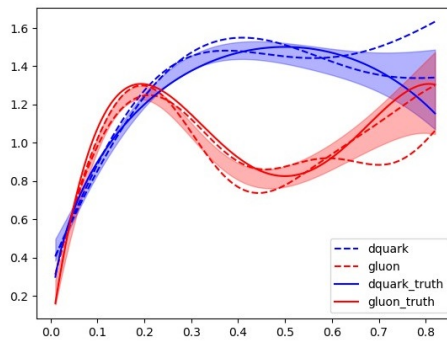
## 6. Rezultati i diskusija

Pri optimizaciji parametra  $\mathbf{w}$  gradijentnim spustom na funkciji gubitka treba pripaziti da parametri koji odgovaraju dostignutom minimumu dobro opisuju skup za učenje pritom ne čineći model presloženim. Potrebno je dakle paziti pri biranju početne točke algoritma te Slika 6 i Slika 8 pomažu u vizualizaciji te problematike. Uz pomoć jednadžbi predviđanja (14)-(16) i pronađenih parametara program generira primjere funkcija diferencijalnog udarnog presjeka na Slici 13 ispod za dva procesa raspršenja od interesa. Dobivena prilagodba jako dobro replicira istinu o diferencijalnim udarnim presjecima za oba procesa raspršenja.



Slika 13: Izgenerirani primjeri funkcija diferencijalnog udarnog presjeka sa pojasom  $\pm 1\sigma$  nakon optimizacije parametara funkcije kovarijance na simulirane podatke

Nadalje, nakon generiranja funkcija diferencijalnih udarnih presjeka izvršena je ekstrakcija partonskih distribucijskih funkcija rješavanjem sustava od dviju jednadžbi s dvije nepoznanice. Na Slici 14 ispod dobiveni rezultati uspoređeni su sa zadanim partonskim distribucijskim funkcijama te je time jasno dočarana uspješnost proučavane metode. Optimizacija parametara gradijentnim spustom, prilagodba na simulirane podatke kao i sama ekstrakcija obavljene su uz pomoć `lsqfitgp` paketa.



Slika 14: Izgenerirani primjeri ekstrahiranih partonskih distribucijskih funkcija sa pojasom  $\pm 1\sigma$

Jedna od prednosti ovakve metode ekstrakcije je izostanak modeliranja partonskih distribucijskih funkcija kao nekog poznatog skupa funkcija, na primjer polinoma. S druge strane, jedna od mana je što se u slučaju modeliranja procesa kao gausijanskog odabire funkcija kovarijance te se tim izborom uvodi određena krutost u cijelu metodu. Druga uočljiva mana je subjektivnost pri izboru parametara funkcije kovarijance kod optimizacije. Naime, u stvarnoj situaciji istraživaču nije dostupna informacija o pravoj ovisnosti diferencijalnog udarnog presjeka te on sam mora procijeniti jeli optimizirani parametri odabrane funkcije kovarijance dobro opisuju podatke. Drugim riječima, istraživač mora prosuditi treba li možda tražiti drugi minimum funkcije gubitka. Jedan od načina da se posljednja mana potencijalno riješi je uvođenje spomenute regularizacije gdje se tada gradijentnim spustom nastoji minimizirati transformiranu funkciju gubitka: funkcija gubitka  $\leftarrow$  funkcija gubitka + složenost modela. Opisanom regularizacijom izbjegla bi se mogućnost prenaučivosti. Globalni minimum regularizirane funkcije gubitka mogao bi se tada tražiti nekim od naprednijih

algoritama minimizacije. Nažalost, `lsqfitgp` Python paket do trenutka pisanja ovog seminarskog rada ne nudi opciju regularizacije kao mogućnost.

## 7. Zaključak

Ispitana je metoda ekstrakcije partonskih distribucijskih funkcija iz podataka diferencijalnog udarnog presjeka za više različitih procesa raspršenja nad istim hadronom. Proučavana metoda bazirana je na specifičnoj metodi regresije gausijanskih procesa. Problem se pojednostavio smanjenjem broja kinematičkih varijabli na jednu te zadržavanjem na samo dvije partonske distribucijske funkcije o kojima ovisi diferencijalni udarni presjek. U programskom jeziku Python, zadavanjem partonskih distribucijskih funkcija koje je cilj replicirati, simulirani su eksperimentalni podatci diferencijalnih udarnih presjeka za dva procesa raspršenja. Uz pomoć `lsqfitgp` paketa partonske distribucijske funkcije su modelirane kao gausijanski procesi te njihovom linearnom transformacijom također i diferencijalni udarni presjeci. Specifičnom metodom regresije nad gausijanskim procesima odrađena je prilagodba na podacima diferencijalnih udarnih presjeka. Rješavanjem sustava od dviju jednažbi s dvije nepoznanice ekstrahirane su partonske distribucijske funkcije. Dobiveni rezultati uspoređeni su sa zadanim partonskim distribucijskim funkcijama koje je bilo cilj replicirati. Zaključuje se da je proučavana metoda ekstrakcije obećavajuća. Uz poneke preinake i nadogradnje opisane u radu vidljiv je veliki potencijal za korištenje ovakve metode ekstrakcije u stvarnom istraživanju u fizici elementarnih čestica.

## Literatura

- [1] “lsqfitgp manual.” Dostupno na: <https://gattocrucco.github.io/lsqfitgp/docs/>.
- [2] A. Isaacs, *Oxford Dictionary of Physics*. Oxford University Press, 2000.
- [3] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [4] J. Šnajder, “Bilješke s predavanja - strojno učenje.” Dostupno na: [https://www.fer.unizg.hr/predmet/struce1/materijali#%23!p\\_rep\\_142310!\\_-210848](https://www.fer.unizg.hr/predmet/struce1/materijali#%23!p_rep_142310!_-210848).
- [5] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [6] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined: Foundations, Algorithms and Applications*. Cambridge University Press, 2020.
- [7] C.E.Rasmussen and C.K.I.Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [9] “Scipy.special manual.” Dostupno na: <https://docs.scipy.org/doc/scipy/reference/special.html>.