

# Određivanje neodređenosti hadronske strukture metodama strojnog učenja

Ivan Ćorić  
20. siječnja 2019.

## Sažetak

U ovom seminaru proučava se mogućnost ekstrakcije Comptonских form faktora (CFF) iz umjetnih (*engl. mock*) podataka, kako bismo dobili uvid u kvalitetu istog postupka na realnim, eksperimentalnim podacima. Kako ne bismo unijeli pristranost pri izboru modela, za model CFF funkcija biramo neuronske mreže. Svaki CFF opisujemo neuronskom mrežom te ih učimo na umjetnim podacima. Relativno uspješno uspijemo ekstrahirati samo CFF funkciju koja najviše doprinosi odgovarajućoj opservabli.

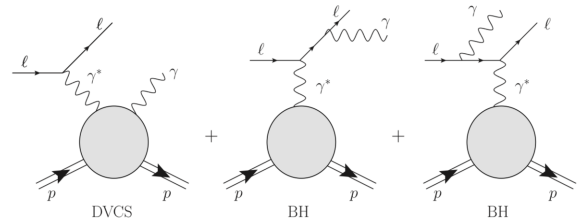
## 1 Uvod

Poznavanjem generaliziranih partonskih distribucija (GPD-ova):  $H^q$ ,  $E^q$ ,  $\tilde{H}^q$ ,  $\tilde{E}^q$ ,  $H_T^q$ ,  $\tilde{H}_T^q$ ,  $E_T^q$ ,  $\tilde{E}_T^q$  dobiva se detaljna slika raspodjele kvarkova i gluona u nukleonu kao što je opisano u [1]. To je dugoročni cilj istraživanja i jedan od razloga konstrukcije novog EIC sudarivača. GPD-ovi su povezani s Comptonским form faktorima (CFF-ovima):  $\mathcal{H}$ ,  $\mathcal{E}$ ,  $\tilde{\mathcal{H}}$ ,  $\tilde{\mathcal{E}}$ ,  $\mathcal{H}_T$ ,  $\mathcal{E}_T$ ,  $\tilde{\mathcal{H}}_T$ ,  $\tilde{\mathcal{E}}_T$ , tzv. faktorizacijskim formulama kao u [2]:

$$\mathcal{F}(\eta, t) = \sum_q e_q^2 \int_{-1}^1 dx \times \left[ \frac{1}{\eta - x - i\epsilon} - \frac{1}{\eta + x - i\epsilon} \right] F^q(x, \eta, t), \quad (1)$$

$$\tilde{\mathcal{F}}(\eta, t) = \sum_q e_q^2 \int_{-1}^1 dx \times \left[ \frac{1}{\eta - x - i\epsilon} + \frac{1}{\eta + x - i\epsilon} \right] \tilde{F}^q(x, \eta, t), \quad (2)$$

gdje je  $\mathcal{F} = \mathcal{H}, \mathcal{E}, \dots$ , a  $F^q = H^q, E^q, \dots$  CFF-ovi su kompleksne funkcije koje su nešto jednostavnije od GPD-ova jer ovise samo o dvije varijable: bezdimenzionalnoj veličini  $\eta$  koju možemo povezati s Bjorkenovim  $x$ -om te Mandelstamovoj  $t$  varijabli.



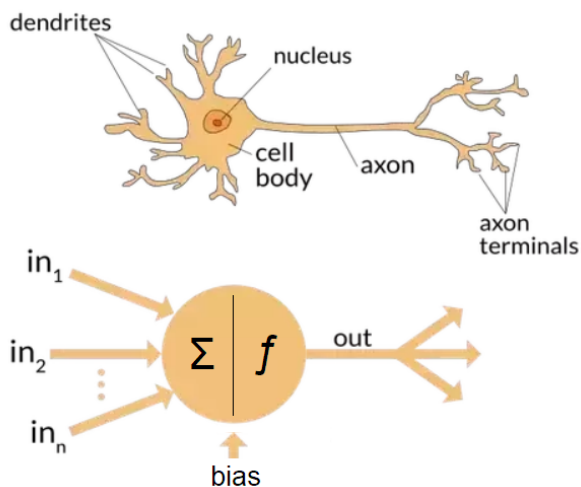
Slika 1: Leptoprodukcija realnog fotona kao koherentna superpozicija DVCS i Bethe-Heitler amplituda. Posuđeno iz [2].

Glavni eksperiment za određivanje CFF funkcija je duboko virtualno komptonско raspršenje (DVCS). U tom eksperimentu ulazni foton je virtualan i dobiva se zračenjem leptona, dakle glavni proces je  $ep \rightarrow ep\gamma$  (tzv. leptoprodukcija fotona). Toj leptoprodukciji doprinosi i druga amplituda, tzv. Bethe-Heitler raspršenje koje možemo vidjeti na slici 1. U eksperimentu se mjeri udarni presjek za  $ep \rightarrow ep\gamma$  i on je dan kao dosta komplicirana funkcija nepoznatih CFF-ova.

Ključno pitanje je kako iz mjerenja DVCS-a odrediti kompleksne CFF funkcije. U najnižem redu doprinose četiri kompleksne CFF funkcije,  $\mathcal{H}$ ,  $\mathcal{E}$ ,  $\tilde{\mathcal{H}}$ ,  $\tilde{\mathcal{E}}$ . U ovom seminaru smo se koncentrirali na situaciju u kojoj imamo raspršenje polariziranog elektrona na nepolariziranom protonu, za što ima dosta mjerenja koja se obično onda izražavaju kao "nepolarizirani" udarni presjek  $\sigma$  (tj. zbroj spin-up i spin-down snopa, *engl. beam spin sum* - BSS) te "polarizirani" udarni presjek  $\Delta\sigma$  (njihova razlika, *engl. beam spin difference* - BSD) te kao njihov omjer  $\sigma_{BSA} = \Delta\sigma/\sigma$  (*engl. beam spin asymmetry* - BSA). Mi u ovom seminaru želimo saznati koliko dobro možemo ekstrahirati CFF funkcije na temelju izmjerenih podataka. Kako ne bismo unijeli pristranost u odabir modela CFF opisujemo neuronskim mrežama.

## 2 Neuronske mreže

Umjetne neuronske mreže motivirane su biološkim neuronskim mrežama te njihovom sposobnošću da nauče uspješno obavljati kompleksne zadatke.



Slika 2: Usporedba biološkog te umjetnog neurona. Slika je posuđena s [3].

Na slici 2 su prikazani biološki te umjetni neuron. Biološki neuron koristi dendrite da prihvati ulazne signale, ulazni signali prolaze kroz tijelo neurona te izlaze kroz akson. Na sličan način umjetni neuron obrađuje podatke. Pretstavimo da je ulazni podatak  $n$ -dimenzionalan. Komponente jednog ulaznog podatka označavamo s  $x_1, x_2, \dots, x_n$ . Svaka komponenta biva pomnožena s odgovarajućom težinom te se tako transformirani ulazni podaci sumiraju te im se još pribroji pristranost (*engl. bias*) kao:

$$u_k = \sum_{j=1}^n w_{kj}x_j + b_k, \quad (3)$$

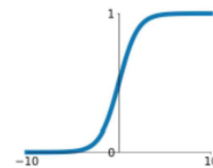
gdje je  $b_k$  još jedan parametar kojeg nazivamo pristranošću (*engl. bias*), a  $k$  odgovara indeksu neurona. Takva suma se propušta kroz nelinearnu funkciju  $f$  kao

$$f_k = f(u_k). \quad (4)$$

Funkcija  $f$  mora biti nelinearna kako bi se mogle modelirati i nelinearne funkcije (jer je linearna kombinacija linearnih kombinacija opet linearna funkcija pa bi mreža mogla opisivati samo linearne funkcije). Za  $f$  se između ostalih najčešće uzimaju tangens hiperbolni, sigmoidalna funkcija, ReLU itd. Te funkcije možemo vidjeti na slici 3.

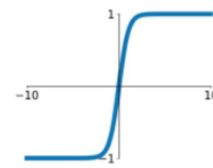
**Sigmoid**

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



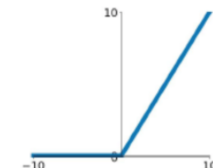
**tanh**

$$\tanh(x)$$

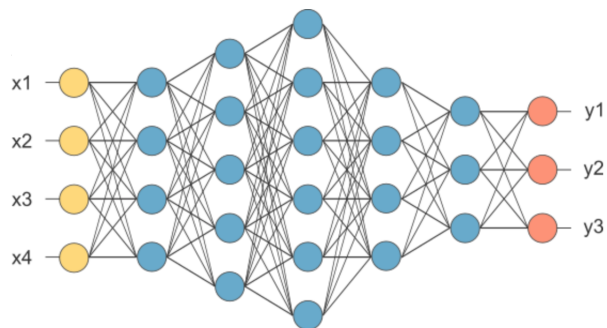


**ReLU**

$$\max(0, x)$$



Slika 3: Prikaz tangensa hiperbolnog, ReLU te sigmoidalne funkcije. Slika posuđena s [4].



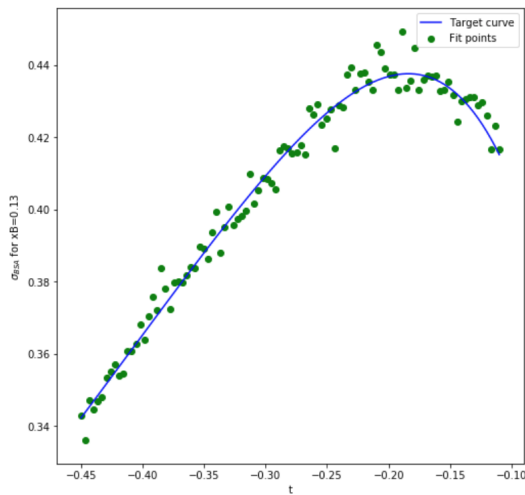
Slika 4: Prikaz neuronske mreže. Slika posuđena s [5].

Neurone nadalje slažemo u slojeve kao što je prikazano na slici 4. Podaci se predaju svakom neuronu u prvom sloju. Izlazi neurona prvog sloja predaju se na ulaz neurona u drugom sloju itd. Brojevi neurona u svakom sloju mogu biti proizvoljni. Ovisno o zadatku koji se rješava (regresija, klasifikacija slika, semantička segmentacija, detekcija objekata, obrada prirodnog jezika (*engl. natural language processing*), ...) arhitektura neuronske može biti značajno različita pa tako imamo konvolucijske neuronske mreže, mreže s povratnom vezom (*engl. recurrent*), ...

Neuronska mreža predstavlja model koji kao parametre ima težine i *bias-e* za svaki neuron. Kako bi model sa što većom točnošću obavljao zadatak koji bismo htjeli, moramo naučiti parametre neuronske mreže. Učenje se odvija algoritmom propagacije unazad (*engl. backpropagation*) čiju ideju ćemo ukratko opisati.

Nakon što ulazni podaci prođu kroz neuronsku mrežu

dobivamo izlaz iz mreže. Prepostavimo da su nam poznati podaci tj. dio znanja koje želimo naučiti našu neuronsku mrežu tj. imamo ulazne i izlazne podatke (takav tip učenja nazivamo nadzirano učenje). Za svaki podatak tada možemo usporediti dobiveni izlaz iz mreže s odgovarajućim izlaznim podatkom. Tada možemo definirati funkciju gubitka (*engl. loss function*) koja nam govori koliko mreža loše odredila izlazne podatke (npr. možemo je definirati kao sumu kvadratnih odstupanja). Cilj je minimizirati tu funkciju (tj. smanjiti grešku). To radimo tako da deriviramo funkciju po svakom parametru, dobijemo gradijente za svaki parametar, te ih promijenimo za taj gradijent pomnožen s nekom malom konstantom (određeno hiperparametrom, tzv. stupnjem učenja (*engl. learning rate*)). Može se pokazati da se dijelovi računa gradijenata koji se dobiju u slojevima bliže izlazu mogu upotrijebiti za računanje gradijenata u nižim slojevima pa se ta operacija može izračunati efikasnije. To je ideja algoritma propagacije unatrag (tj. gradijenti se propagiraju iz viših slojeva u niže).

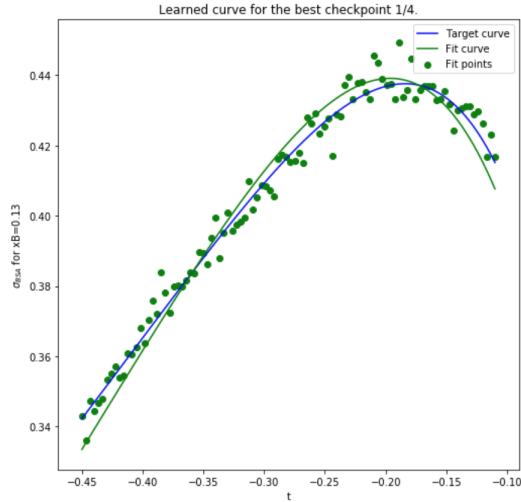


Slika 6: Prikaz mock podataka.

### 3 Rezultati

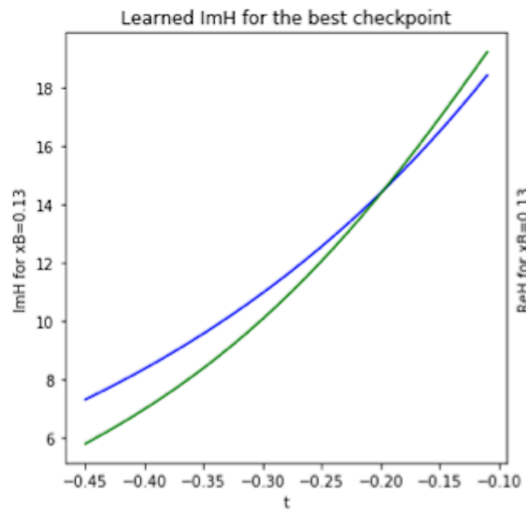
Kako bismo provjerili koliko dobro možemo koristiti neuronske mreže na stvarnim podacima, odlučili smo provesti eksperimente na umjetnim (tzv. mock) podacima. Uzeli smo kao primjer popularni Goloskokov-Kroll (GK) model za modeliranje CFF funkcija. Onda smo koristeći GK model generirali podatke za različite vrijednosti Mandelstamove  $t$  varijable, u te podatke smo unijeli šum i pokušali na tim šumovitim podacima što bolje ekstrahirati CFF funkcije u GK modelu. Zadatak nam je bio probati dobiti što bolje rezultate za ekstrakciju CFF funkcija u jednoj dimenziji (fiksni  $\eta$ ) te u dvije dimenzije. Na slici 5 prikazani su CFF-ovi za GK

model te je na slici 6 prikaz umjetnih (mock) podataka. Za početak radimo eksperimente samo na jednodimenzionalnom skupu za fiksni Bjorkenov  $x$ ,  $x_B = 0.13$ , što fiksira varijablu  $\eta$ , a različite vrijednosti Mandelstamove  $t$  varijable.

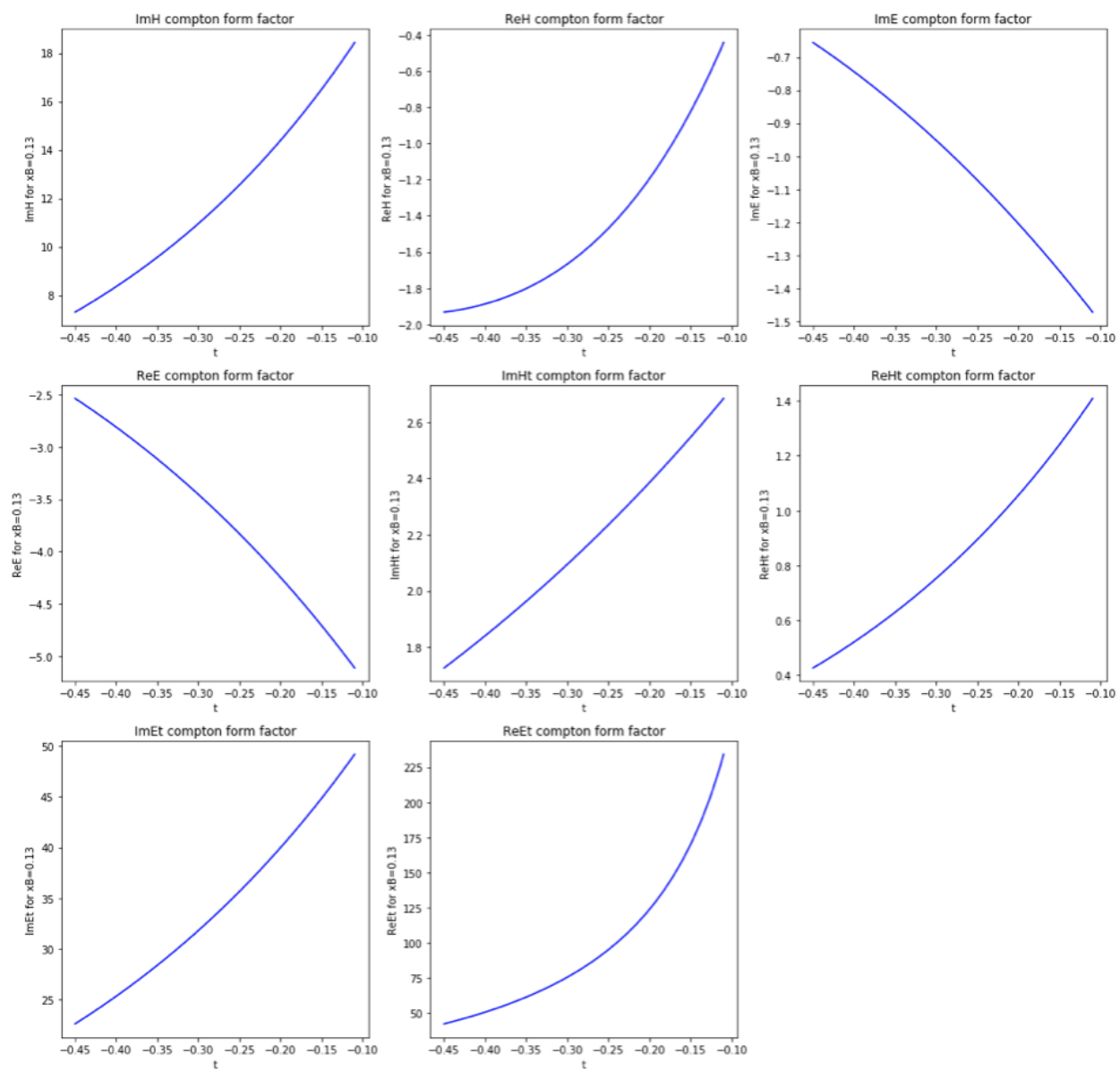


Slika 7: Prikaz naučenog  $\sigma_{BSA}$  gdje je plava krivulja  $\sigma_{BSA}$  u GK modelu dok je zelena naučena krivulja.

Nad umjetnim podacima nadalje radimo uobičajenu, slučajnu podjelu u skup za treniranje i validaciju u omjeru 2 : 1. Tada učimo model na skupu za treniranje te na skupu za validaciju testiramo kako radi na podacima koje još nije vidio tj. koliko dobro generalizira. Kao najbolje rješenje uzimamo model koji postiže najmanju pogrešku na validacijskom skupu. Kao pogrešku smo uzeli uobičajenu pogrešku kvadratnog odstupanja.



Slika 8: Prikaz naučenog  $Im\mathcal{H}$  gdje je plava krivulja  $Im\mathcal{H}$  u GK modelu dok je zelena naučena krivulja.



Slika 5: Prikaz comptonskih form faktora u GK modelu za fiksni  $\eta$ .

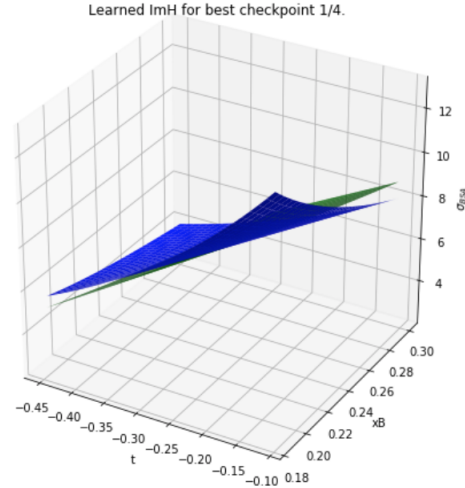
Svaki od CFF-ova opisujemo zasebnom neuronskom mrežom te treniramo mrežu na mnogo različitih načina. Kada bismo sve CFF funkcije trenirali istovremeno, model ne bi naučio niti jedan CFF već bi uspio postići dobro opisivanje  $\sigma_{BSA}$  (dano na slici 7) drugačijom kombinacijom CFF funkcija kao što je prikazano na slici 9.

To nas je potaknulo da prvo pokušamo naučiti najveći doprinos  $\sigma_{BSA}$  što čini imaginarni dio CFF  $\mathcal{H}$  funkcije, dok sve ostale CFF funkcije držimo zamrznutim. Taj pristup se pokazao dobar te smo relativno uspješno naučili imaginarni dio  $\mathcal{H}$  funkcije kao što je prikazano na slici 8.

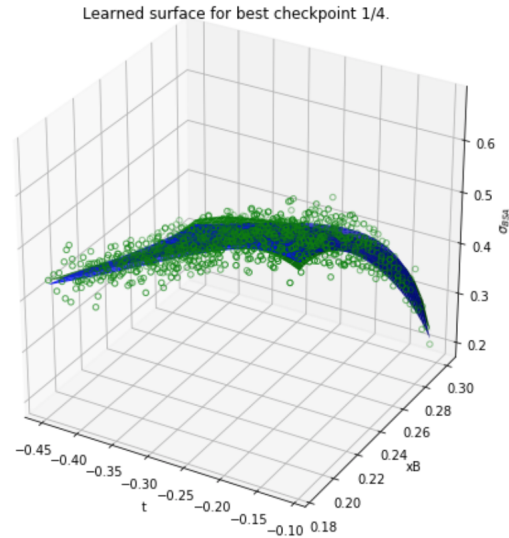
Nadalje spremili smo težine za tako naučeni imaginarni dio  $\mathcal{H}$  funkcije. U sljedećim treninzima smo pokušavali naučiti sljedeći ili sljedeća dva doprinosa  $\sigma_{BSA}$ . Pokrenuli smo stotine eksperimenata s mnoštvom različitih hiperparametara. Pokušavali smo mijenjati dubinu mreže za sljedeća dva doprinosa, no to nije dovelo do učenja istih. Kako je  $\sigma_{BSA}$  već relativno uspješno naučena, s naučenim glavnim doprinosom ( $Im\mathcal{H}$ ) gradijenti koji se propagiraju unatrag do mreža sljedećih doprinosa su maleni. Uz to, sljedeća dva doprinosa su za red do dva veličine (kao u [6]) manja od glavnog ( $Im\mathcal{H}$ ) doprinosa te gradijenti bivaju još više umanjeni. Onda smo pokušali provoditi eksperimente s različitim *learning rate*-ovima koji bi se kretali između nekoliko redova veličine (npr. od  $10^{-2}$  do  $10^8$ ) za svakog od sljedeća dva doprinosa nakon  $Im\mathcal{H}$  (tj. radili smo pretragu po rešetci za *learning rate* hiperparametar (*engl. grid search*)), no niti to nije pomoglo, jedino što smo uspjeli naučiti je pristranost za imaginarni dio  $\mathcal{E}$  CFF funkcije kao što možemo vidjeti na slici 10.

Nakon tih eksperimenata pokušavali smo početi trenirati druga dva doprinosa, ali ovaj put prije nego se  $Im\mathcal{H}$  potpuno naučio kako bi gradijenti bili veći za preostale funkcije, no niti tako nismo uspjeli naučiti sljedeća dva doprinosa.

Nakon treniranja modela u jednoj dimenziji odlučili smo pokušati vidjeti kako model funkcionira u dvije dimenzije tj. za proizvoljne  $\eta$  (odnosno Bjorkenov  $x$ ) te Mandelstamovu varijablu  $t$ . Provodili smo iste tipove eksperimenata koji ovoga puta trajali znatno duže zbog toga što ima mnogo više podataka. Nakon prvog tipa treninga uspjeli smo dobiti relativno dobar  $Im\mathcal{H}$  kao što možemo vidjeti na slici 11, pripadni  $\sigma_{BSA}$  je dan na slici 12.

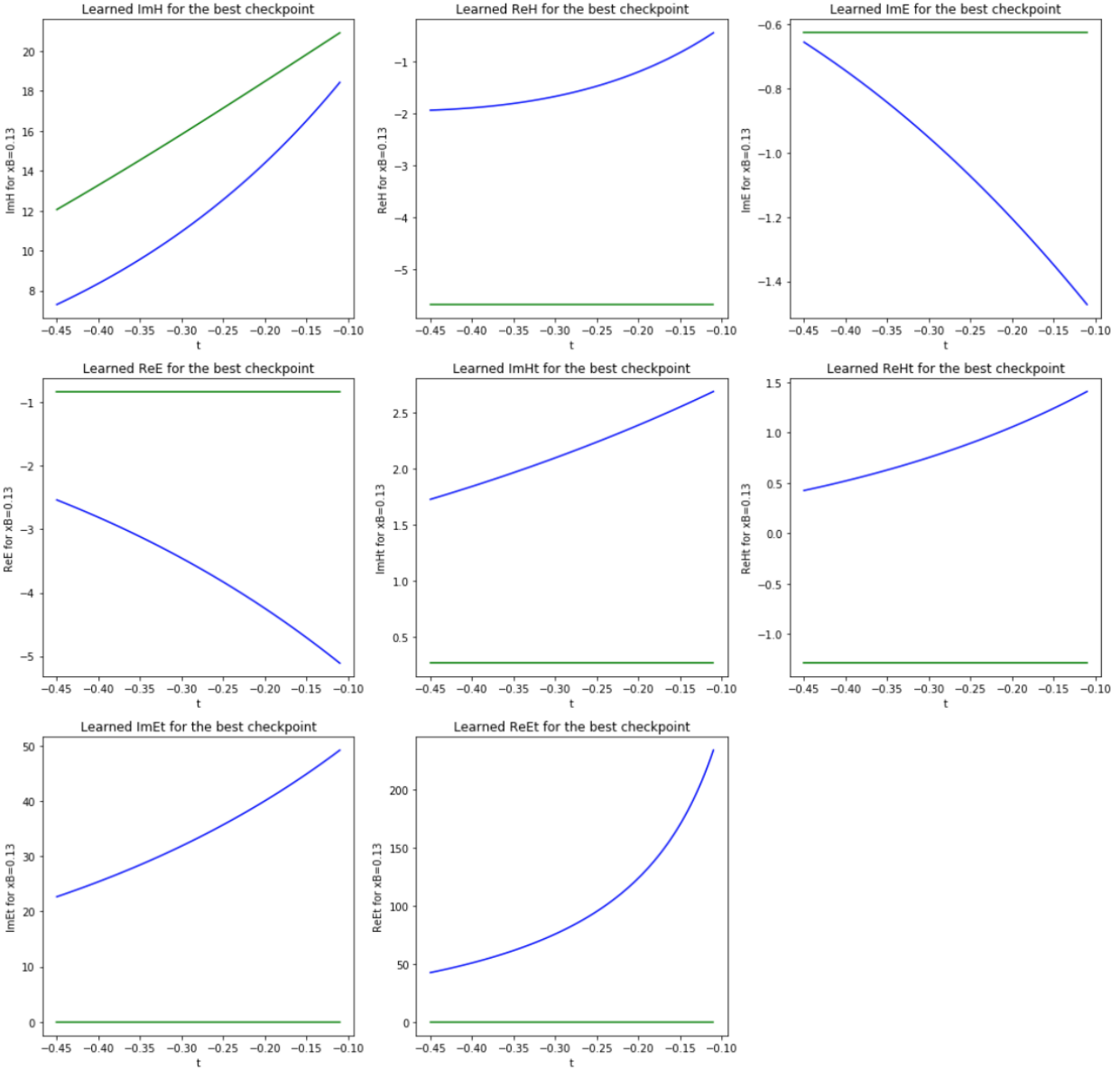


Slika 11: Prikaz naučene  $Im\mathcal{H}$  funkcije (zelena ploha) te odgovarajuće funkcije u GK modelu (plava ploha).

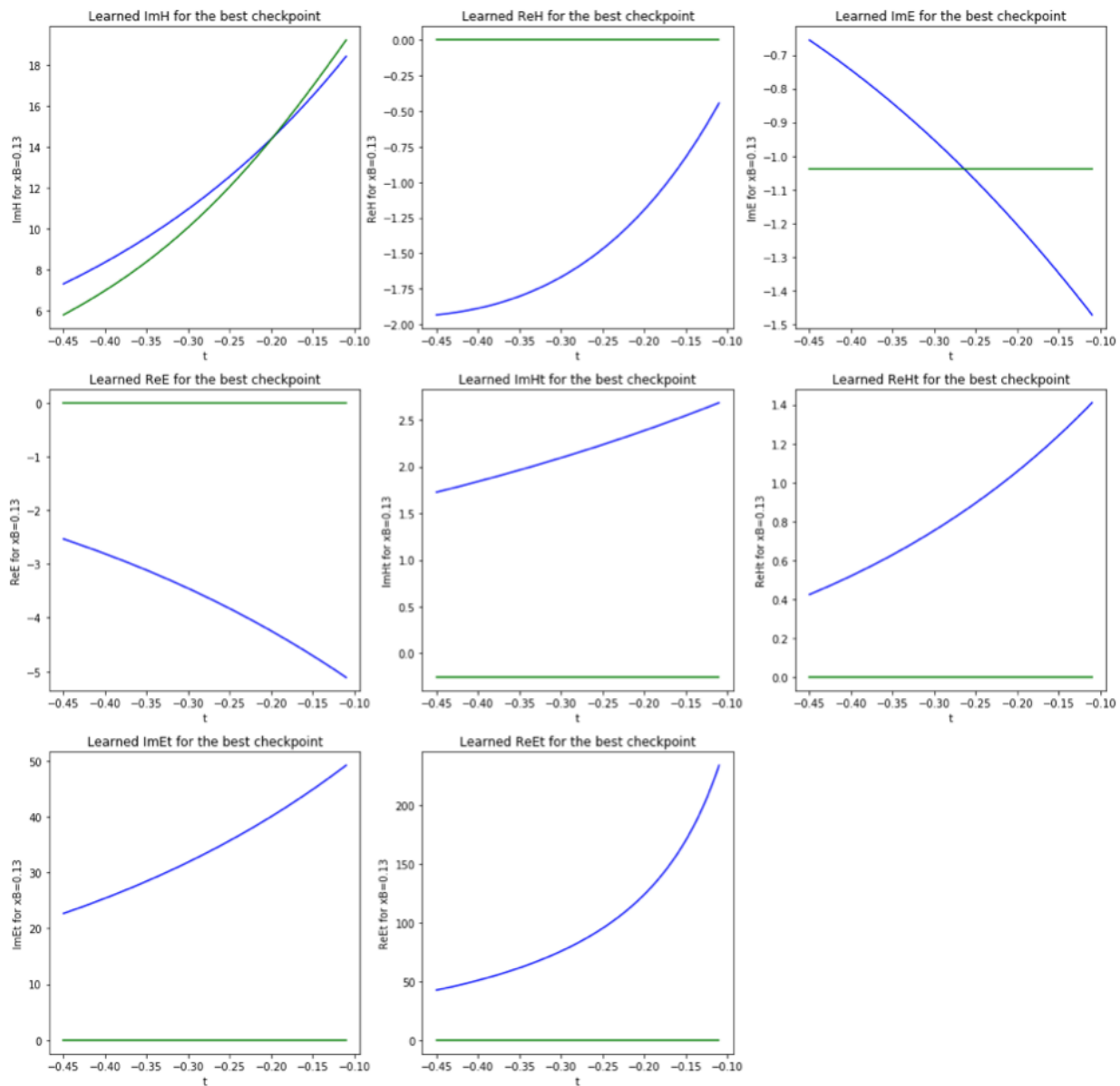


Slika 12: Prikaz naučene  $\sigma_{BSA}$  te odgovarajuće funkcije u GK modelu.

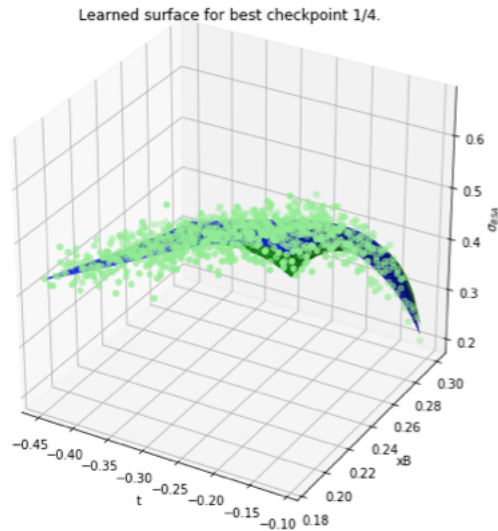
Nakon treniranja samo  $Im\mathcal{H}$ , kao i u jednodimenzionalnom slučaju počinjemo trenirati i ostale CFF funkcije, no ovdje smo odlučili od samog početka zanemariti sve doprinose osim  $Im\tilde{\mathcal{H}}$  i  $Im\mathcal{E}$ , koji nakon  $Im\mathcal{H}$  najviše doprinose (vidi [6]), kako brzina izvođenja ne bi bila pre-mala. Dobiveni rezultati dani su na slikama 13 i 14.



Slika 9: Prikaz naučenih comptonskih form faktora gdje su plavom bojom označene CFF funkcije u GK modelu, a zelenom bojom naučene CFF funkcije.



Slika 10: Prikaz naučenih comptonskih form faktora gdje su plavom bojom označene CFF funkcije u GK modelu, a zelenom bojom naučene CFF funkcije.



Slika 13: Prikaz naučene  $\sigma_{BSA}$  te odgovarajuće funkcije u GK modelu.

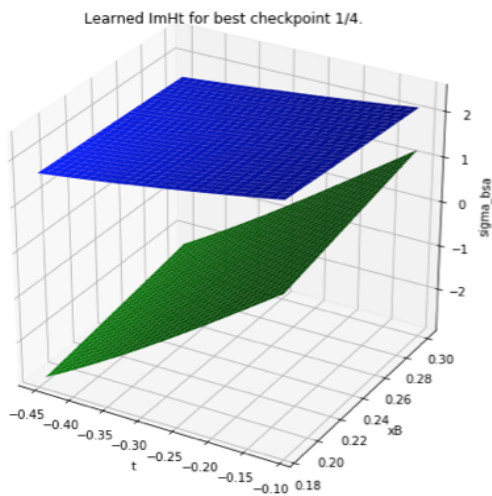
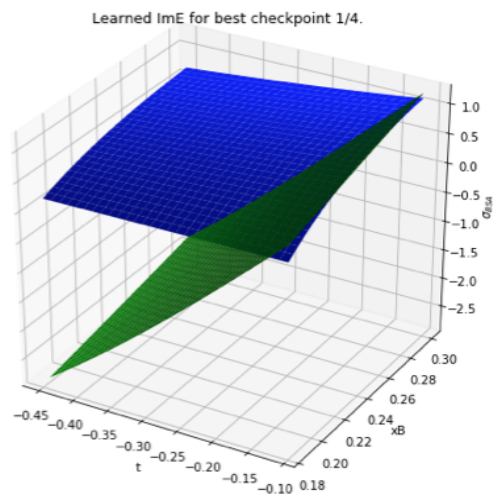
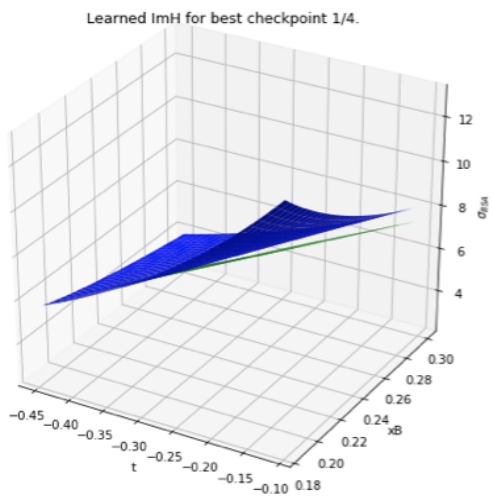
## 4 Zaključak

U ovom seminaru proučavali smo koliko dobro možemo ekstrahirati CFF funkcije pomoću neuronskih mreža iz umjetnih podataka, tako da bolje razumijemo nesigurnost tog postupka kada ga radimo na realnim eksperimentalnim podacima. Iscrpnim eksperimentiranjem, za jednodimenzionalne (fiksni  $\eta$ ), kao i za dvodimenzionalne podatke uspjeli smo relativno dobro ekstrahirati samo CFF funkciju koja najviše doprinosi  $\sigma_{BSA}$ .

## Literatura

- [1] A. Accardi, J. L. Albacete, M. Anselmino, N. Armesto, E. C. Aschenauer, A. Bacchetta, D. Boer, W. K. Brooks, T. Burton, N. B. Chang, W. T. Deng, A. Deshpande, M. Diehl, A. Dumitru, R. Dupré, R. Ent, S. Fazio, H. Gao, V. Guzey, H. Hakobyan, Y. Hao, D. Hasch, R. Holt, T. Horn, M. Huang, A. Hutton, C. Hyde, J. Jalilian-Marian, S. Klein, B. Kopeliovich, Y. Kovchegov, K. Kumar, K. Kumerički, M. A. C. Lamont, T. Lappi, J. H. Lee, Y. Lee, E. M. Levin, F. L. Lin, V. Litvinenko, T. W. Ludlam, C. Marquet, Z. E. Meziani, R. McKeown, A. Metz, R. Milner, V. S. Morozov, A. H. Mueller, B. Müller, D. Müller, P. Nadel-Turonski, H. Paukunen, A. Prokudin, V. Ptitsyn, X. Qian, J. W. Qiu, M. Ramsey-Musolf, T. Roser, F. Sabatié, R. Sassot, G. Schnell, P. Schweitzer, E. Sichtermann, M. Stratmann, M. Strikman, M. Sullivan, S. Taneja, T. Toll, D. Trbojevic, T. Ullrich, R. Venugopalan, S. Vigdor, W. Vogelsang, C. Weiss, B. W. Xiao, F. Yuan, Y. H. Zhang, and L. Zheng. Electron Ion Collider: The Next QCD Frontier - Understanding the glue that binds us all. *arXiv e-prints*, page arXiv:1212.1701, December 2012.
- [2] Krešimir Kumerički, Simonetta Liuti, and Hervé Moutarde. GPD phenomenology and DVCS fitting. Entering the high-precision era. *European Physical Journal A*, 52:157, June 2016.
- [3] Nagyfi Richárd. The differences between artificial and biological neural networks. <https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7>, 2018. [Online; accessed 18-January-2019].
- [4] Activation functions. [https://www.julyedu.com/question/big/kp\\_id/26/ques\\_id/1044](https://www.julyedu.com/question/big/kp_id/26/ques_id/1044). [Online; accessed 18-January-2019].
- [5] Peter Kottas. <https://www.quora.com/How-does-deep-learning-work-and-how-is-it-different-from-normal-neural-networks-applied-with-SVM-How-does-one-go-about-starting-to-understand-them-papers-blogs-articles>. [Online; accessed 18-January-2019].
- [6] P. Kroll, H. Moutarde, and F. Sabatié. From hard exclusive meson electroproduction to deeply virtual Compton scattering. *European Physical Journal C*, 73:2278, January 2013.





Slika 14: Prikaz naučenih comptonskih form faktora gdje su plavom bojom označene CFF funkcije u GK modelu, a zelenom bojom naučene CFF funkcije.