

Testovanie hyperparametrov neurónových sietí pomocou PyTorch

Dmytro Skrypchenko

Abstract

Tento projekt je zameraný na testovanie vplyvu rôznych hyperparametrov, ako sú topológia siete, optimalizátor, learning rate a aktivačná funkcia, na presnosť modelu. Hyperparametre sa optimalizujú pre konkrétny dataset, ktorého popis bude uvedený nižšie.

I. DATASET

Ladenie a hľadanie optimálnych parametrov prebiehalo pre dataset "Diabetes Dataset", ktorý obsahuje informácie o ľuďoch, ich zdravotných charakteristikách a údaj o tom, či má človek cukrovkou alebo nie. Dataset obsahuje 8 atribútov a 1 cieľovú premennú[1].

Atribút	Priemerná hodnota	Štandardná odchýlka	Minimálna hodnota	Maximálna hodnota	Medián
Pregnancies	3.8451	3.3696	0.0000	17.0000	3.0000
Glucose	120.8945	31.9726	0.0000	199.0000	117.0000
BloodPressure	69.1055	19.3558	0.0000	122.0000	72.0000
SkinThickness	20.5365	15.9522	0.0000	99.0000	23.0000
Insulin	79.7995	115.2440	0.0000	846.0000	30.5000
BMI	31.9926	7.8842	0.0000	67.1000	32.0000
DiabetesPedigreeFunction	0.4719	0.3313	0.0780	2.4200	0.3725
Age	33.2409	11.7602	21.0000	81.0000	29.0000

TABLE I: Štatistické charakteristiky datasetu

Atributy zahŕňajú faktory ako vek, hladinu glukózy, krvný tlak, hrúbku kože, hladinu inzulínu, index telesnej hmotnosti (BMI), genetickú predispozíciu a počet tehotenstiev. Každý z týchto atribútov poskytuje dôležité informácie, ktoré môžu ovplyvniť výskyt cukrovky. Cieľovou premennou je, či daná osoba trpí cukrovkou (1) alebo nie (0), čo umožňuje analyzovať vzťah medzi týmito zdravotnými ukazovateľmi a výskytom cukrovky.

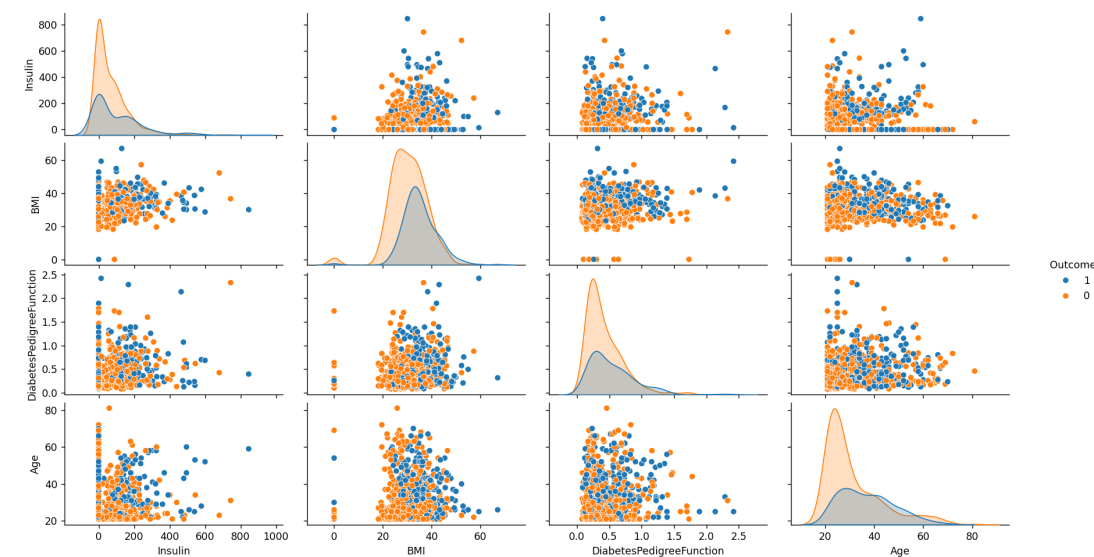


Fig. 1: Vzťahy medzi atribútmi 1 časť

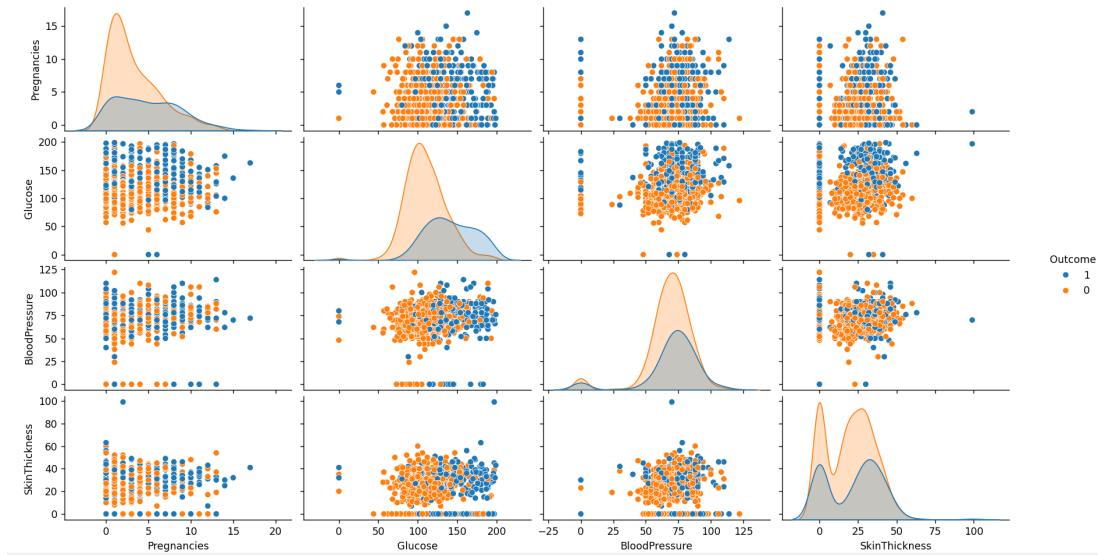


Fig. 2: Vzťahy medzi atribútmi 2 časť

Z grafov je zrejmé, že triedy sú pomerne silno zamiešané. Najlepšie a najviac oddelené triedy sa javia na základe atribútov Glucose, BloodPressure a SkinThickness. Očakávať teda vysokú presnosť modelu na tomto datasete bude výzvou, pretože separácia medzi triedami nie je výrazná.

II. TOPOLOGIA

Základným modelom na testovanie vplyvu topológie je klasický viacvrstvový perceptron. Počas testovania boli všetky ostatné hyperparametre nastavené rovnako. Rovnako dôležitú úlohu zohráva inicializácia počiatočných parametrov neurónovej siete, ktorá bola zabezpečená použitím rovnakého random seed pre každý model.

Hyperparameter	Value
Epochs	200
Learning Rate	0.003
Batch Size	8
Random Seed	12
Training Example Fraction	0.8
Optimizer	Adam
Activation Function	ReLU

TABLE II: Ostatné hyperparametre

Pri týchto hyperparametroch bolo testovaných 6 rôznych topológií, ktoré boli porovnávané na základe metriky presnosti (accuracy)

Počet skrytých vrstiev	Počet neurónov v skrytých vrstvách	Presnosť (Accuracy)
1	8	0.8117
1	16	0.8117
2	8	0.8182
2	16	0.7468
3	8	0.7662
3	16	0.7338

TABLE III: Porovnanie topológií na základe presnosti

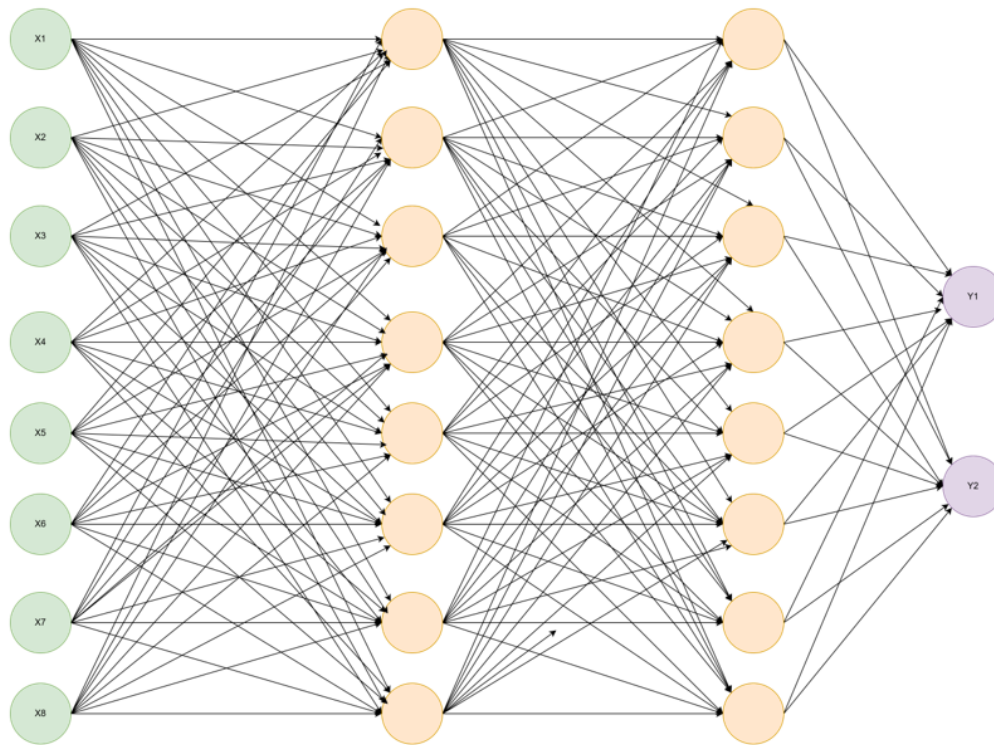


Fig. 3: Najlepšia topológia

Teda vidíme, že najlepšiu presnosť dosiahla stredne zložitá topológia. Menej zložené topológie tiež ukázali vysokú presnosť, ale pri zložitejších topológiách sa presnosť viditeľne znížila, čo naznačuje možný overfitting.

III. OPTIMALIZÁTOR

Ďalším hyperparametrom na optimalizáciu je optimalizátor. Pri tom budeme používať najlepšiu topológiu, ktorú sme určili v predchádzajúcom kroku, a ostatné hyperparametre necháme rovnaké, len samozrejme budeme meniť optimalizátor.

Optimalizátor	Presnosť (Accuracy)
Adam	0.8182
SGD	0.7727
RMSprop	0.7922

TABLE IV: Porovnanie optimalizátorov na základe presnosti

Adam dosiahol najlepšiu presnosť, pretože kombinuje výhody adaptívnych momentových metód a rýchlej konvergenencie, čo vedie k stabilnejšiemu a efektívnejšiemu trénovaniu modelu v porovnaní s ostatnými optimalizátormi.

IV. UČIACI PARAMETER

Pri testovaní rôznych učiacich parametrov budeme používať najlepšiu topológiu a optimalizátor, pričom ostatné hyperparametre zostanú nezmenené.

Learning rate	Presnosť (Accuracy)
0.01	0.8182
0.001	0.7792
0.003	0.8182
0.005	0.7987
0.05	0.6429

TABLE V: Porovnanie učiacich parametrov na základe presnosti

Vidíme, že najlepšiu presnosť model dosiahol pri learning rate 0.01 a 0.003. Vysokú presnosť model ukázal aj pri learning rate 0.001 a 0.005, zatiaľ čo pri learning rate 0.05 model dosiahol najhoršiu presnosť. Tento výsledok naznačuje, že vysoká hodnota učenia môže viesť k overfittingu.

V. AKTIVAČNÁ FUNKCIA

Optimálnu aktivačnú funkciu budeme hľadať s použitím najlepšej topológie, optimalizátora a learning rate 0.01. Pre porovnanie vyberieme 5 najpopulárnejších aktivačných funkcií.

Názov aktivačnej funkcie	Vzorec	Presnosť (Accuracy)
ReLU	$f(x) = \max(0, x)$	0.8182
SELU	$f(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases}$	0.8312
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	0.6818
LeakyReLU	$f(x) = \max(\alpha x, x), \quad \alpha = 0.01$	0.8052
Tanh	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	0.6429

TABLE VI: Porovnanie aktivačných funkcií parametrov na základe presnosti

Najlepšiu presnosť dosiahla aktivačná funkcia SELU, dobré výsledky dosiahli aj ReLU a LeakyReLU, zatiaľ čo Sigmoid a Tanh mali výrazne horšiu presnosť. Tieto výsledky naznačujú, že pre konkrétny dataset použitie neoptimálnej aktivačnej funkcie môže viesť k výrazne horším výsledkom. Avšak to neznamená, že jedna aktivačná funkcia je vždy horšia a iná lepšia, pretože stále záleží na dátach.

VI. CONSULUSION

Pomocou experimentov sme zistili, že pre tento dataset je optimálna topológia, ktorá má dve skryté vrstvy s počtom neurónov 8. Najvhodnejší optimalizátor je Adam, optimálna hodnota učebného parametra je 0.01 a najlepšia aktivačná funkcia je SELU.

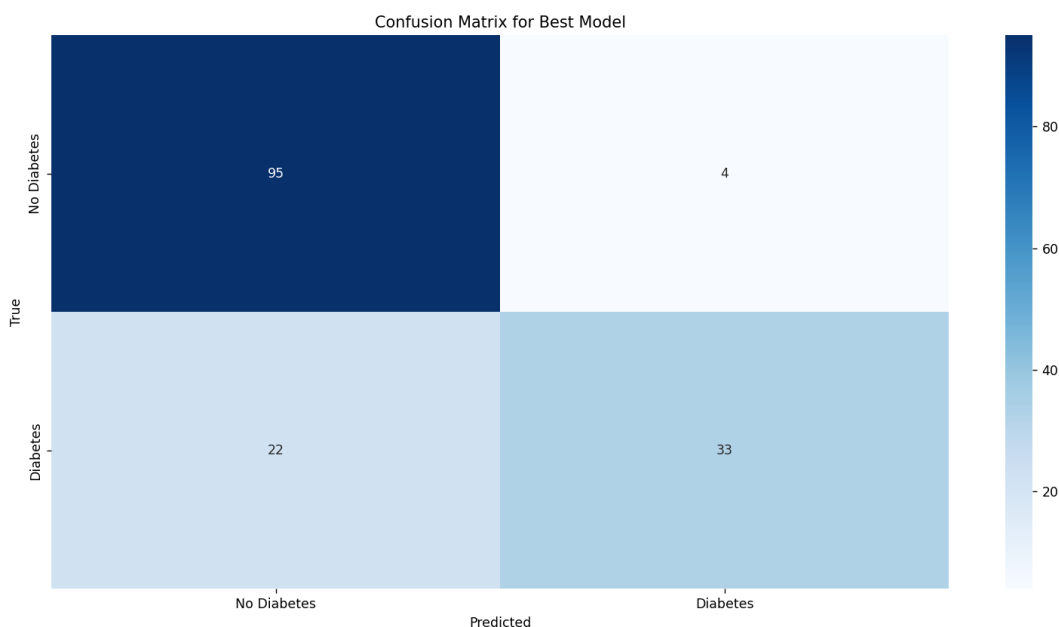


Fig. 4: Konfúzna matica pre najlepší dosiahnutý model

Hlavným výsledkom týchto experimentov je, že použitím suboptimálnych alebo optimálnych hyperparametrov je možné výrazne zvýšiť výkonnosť a presnosť modelu. Kľúčové je, že všetko závisí od dát, pretože v prípade iného datasetu by najlepšia konfigurácia mohla vyzeráť úplne inak.

REFERENCES

- [1] Kaggle. Diabetes Dataset. Dostupné na: <https://www.kaggle.com/datasets/krishu22/diabetes-dataset>