



## SFC - technická správa projektu

**Martin Koči** (xkocim05)

# Obsah

<b>1</b>	<b>Zadanie</b>	<b>2</b>
<b>2</b>	<b>Teória</b>	<b>2</b>
2.1	Konvolučná vrstva . . . . .	2
2.2	Max pooling . . . . .	2
2.3	Average pooling . . . . .	2
2.4	Plno prepojená vrstva . . . . .	3
2.5	Sigmoida . . . . .	3
<b>3</b>	<b>Implementácia</b>	<b>3</b>
<b>4</b>	<b>Manuál</b>	<b>3</b>
4.1	Požiadavky na spustenie . . . . .	3
4.2	Spustenie programu . . . . .	3
4.3	Nastaviteľné parametre . . . . .	3

# 1 Zadanie

Zadaním projektu bolo implementovať konvolučnú neurónovú sieť a demonštrovať jej činnosť. Pre tento účel bola implementovaná konzolová aplikácia, ktorá obsahuje implementovanú konvolučnú vrstvu, max pooling vrstvu, average pooling vrstvu, plne prepojenú vrstvu a aktivačnú funkciu sigmoidu. Z jednotlivých vrstiev bola vytvorená jednoduchá konvolučná neurónová sieť, ktorá sa učí pomocou metódy backpropagation. Chybová funkcia bola použitá cross-entropy loss.

## 2 Teória

Táto kapitola obsahuje stručnú teóriu k jednotlivým vrstvám konvolučnej neurónovej siete. Začínajúc konvolučnou vrstvou. Teória bola čerpaná z [2] a [1].

### 2.1 Konvolučná vrstva

Konvolučná vrstva vykonáva operáciu konvolúcie nad vstupným vektorom. Skladá sa z učiacich sa filtrov. Počet filtrov býva väčšinou desiatky až stovky. Veľkosť filtrov väčšinou býva  $2 \times 2$  alebo  $3 \times 3$ , ale nič nebráni nastaviť väčšiu alebo menšiu hodnotu pokiaľ je kladná. Filtre sa postupne posúvajú po vstupom vektore a vytvárajú výstupný vektor hodnôt. Ak má vstupný vektor viacej dimenzií tak sa vo výstupnom vektore pre danú pozíciu všetky hodnoty z filtra cez jednotlivé dimenzie sčítajú. Dimenzia výstupného vektora je daná počtom filtrov. Konvolučná vrstva okrem nastavenia počtu filtrov a veľkosti filtra má ešte 2 nastaviťelné parametre a to posun a výplň. Posun nastavuje o koľko prvkov sa má filter posunúť čím väčšie číslo tým je výsledná šírka a výška výstupného vektora menšia. Výplň obvykle nulová výplň. Výplní okolie vstupného vektora nulami. Čím väčšia je veľkosť výplne tým väčší je výsledný vektor.

### 2.2 Max pooling

Obecne pooling vrstvy zoskupujú (zmenšujú) vstupný vektor. Skladajú sa z filtra ktorý postupne prechádza vstupný vektor a vykonáva dopredu určenú funkciu, ktorej výsledok sa zapisuje do výstupného zmenšeného vektora. Obvykle sa vstupný vektor zmenšuje o polovicu. Nastaviťelné parametre sú veľkosť filtra prípadne posun filtra v vstupnom vektore. Obvykle sa filter nastavuje na veľkosť  $2 \times 2$  a veľkosť posunu na 2. Konkrétne pri max pooling sa vykonáva operácia maximum nad oblasťou ktorú filter prechádza vo vstupnom vektore. Pri vykonávaní operácie sa môžu zistiť aj indexy maximálnych hodnôt pre jednoduchšie šírenie gradientu pri backpropagation. Pri backpropagation sa gradient propaguje len do indexov vektora kde sa nachádzali maximálne hodnoty pri doprednom prechode sieťou a ostatné hodnoty zostávajú na nule.

### 2.3 Average pooling

Average pooling funguje rovnako ako max pooling až na to že vykonáva operáciu priemeru nad danou oblasťou vstupného vektora danou filtrom. Pri backpropagation rozdeľuje gradient rovnomerne do oblasti ktoré spriemeroval.

## 2.4 Plno prepojená vrstva

Plne prepojená vrstva sa skladá z  $n$  vstupných a  $m$  výstupných neurónov. Ďalej obsahuje vektor váh a vektor biasov. Pri tejto sieti sa vynásobí vektor váh so vstupným vektorom a pričíta sa vektor biasov.

## 2.5 Sigmoida

Aktivačná funkcia sigmoida normalizuje vstupné hodnoty vektora do rozsahu  $0 - 1$ .

# 3 Implementácia

Projekt bol implementovaný v programovacom jazyku Python za pomoci knižnice `numpy`. Implementácia sa skladá z 3 súborov: `cnn.py`, `utils.py`, `Layers.py`. V súbore `Layers.py` sa nachádzajú implementované triedy pre jednotlivé vrstvy konvolučnej neurónovej siete. Každá trieda obsahuje dve metódy. Prvá metóda má názov `forward`, táto metóda berie ako vstupný parameter vstupný vektor a vráti výsledok doprednej operácie danej vrstvy. Druhá metóda s názvom `backward` implementuje spätný prechod danou vrstvou siete. Na vstupe má gradient s predchádzajúcej vrstvy, aktualizuje učiace sa parametre vrstvy a vráti nový gradient na výstup ktorý putuje ďalej sieťou. Následne súbor `utils.py` obsahuje dodatočné funkcie na počítanie cross-entropy loss, vypisovanie matíc, vypisovanie operácii nad maticami, funkciu na učenie siete a predikciu a tak ďalej. Následne posledný súbor `cnn.py` obsahuje načítanie parametrov pre konvolučnú neurónovú sieť a zahajuje proces učenia a samotnú demonštráciu funkčnosti konvolučnej neurónovej siete. Sieť je inicializovaná do listu ktorý sa pri doprednom prechode postupne prechádza a volá sa funkcia `forward` po doprednom prechode sa spočíta chyba a jej gradient a postupne sa začne šíriť spätným prechodom siete pri ktorom sa list prechádza v opačnom poradí a volá sa funkcia `backward`.

# 4 Manuál

V tejto kapitole sa nachádza stručný manuál k práci s dostupnými zdrojovými súbormi.

## 4.1 Požiadavky na spustenie

Pre správne fungovanie programu je požadovaná aspoň minimálna verzia pythonu 3.8 a knižnica `numpy` s verziou minimálne 1.24.4.

## 4.2 Spustenie programu

Program sa spustí jednoducho v konzoly s príkazom:

```
python3.8 cnn.py
```

## 4.3 Nastaviteľné parametre

Program má niekoľko nastaviteľných parametrov, ktoré sa nastavujú ale až pri spustení programu. Po spustení programu sa program spýta na jednotlivé nastavenie vstupných parametrov a následne začne vykonávať demonštráciu. Nastaviteľné parametre sú nasledujúce:

- Veľkosť filtra konvolučnej vrstvy (predvolená hodnota 3)
- Posun filtra konvolučnej vrstvy (predvolená hodnota 1)
- Výplň vstupného vektoru konvolučnej vrstvy (predvolená hodnota 0)
- Počet filtrov konvolučnej vrstvy (predvolená hodnota 1)
- Použitie max pooling vrstvy (1) alebo average pooling vrstvy (0) - predvolená hodnota 1
- Počet epoch (predvolená hodnota 100)
- Nastavenie learning ratu (predvolená hodnota 0.1)
- Nastavenie výpisu chyby po jednej epoche (predvolená hodnota áno)
- Nastavenie výpisu demonštrácie učenia (predvolená hodnota áno)
- Možnosť nastavenie dimenzie vstupu (predvolená hodnota 1)
- Možnosť nastavenia šírky a výšky vstupu (je minimálna na základe nastavení siete, a predvolená je o 2 väčšie ako minimálna).
- Nastavenie vstupnej matice a jej očakávaného výstupu. Nastaviť sa dá len jeden vstup a učí sa len s jedným vstupom pre prakticky použitie tejto aplikácie na demonštráciu činnosti.

Počas učenia sa objavujú dve hlášky `Ďalší krok [stlačte ľubovoľné tlačítko]` a `Vynechať vypisovanie výpočtu pre n epocho [default 0] : .` Pričom prvá hláška po stlačení ľubovoľnej klávesy pokračuje na ďalší krok učenia. Druhá hláška sa zobrazí vždy po ukončení jednej epochy a navrhne preskočenie výpisu demonštrácie učenia na n epoch počas ktorých sa bude vypisovať len chyba siete. Pre ukončenie programu stačí napísať buď viac epoch ako je nastavených ako maximum alebo stačí stlačiť (CTRL + C).

## Referencie

- [1] BURKOV, A. *The Hundred-Page Machine Learning Book*. Quebec, Canada: Andriy Burkov, 2019. ISBN 978-1-9995795-0-0.
- [2] JAMES, G., WITTEN, D., HASTIE, T. a TIBSHIRANI, R. *An Introduction to Statistical Learning: with Applications in R*. 1. vyd. New York, NY: Springer, 2013. Springer Texts in Statistics. ISBN 978-1-4614-7137-0.