

Zadanie projektu – Umelé neurónové siete

Úlohy:

5a. Klasifikácia do skupín pomocou viacvrstvovej perceptrónovej (MLP) siete. (1,5 b)

5b. Aproximácia nelineárnej funkcie pomocou MLP siete. (1,5 b)

6. Klasifikácia na medicínskych dátach pomocou MLP siete. (4 b)

7. Rozpoznávanie rukou písaných číslíc pomocou MLP siete. (4 b)

8. Rozpoznávanie rukou písaných číslíc pomocou konvolučnej neurónovej siete. (6 b)

Úlohy 5 predstavujú jednoduchšie príklady na aproximáciu a klasifikáciu pomocou viacvrstvovej perceptrónovej siete. Vašou úlohou je modifikovať demo ukážky k prednáške 6 (2. prednáška NS) podľa zadáných úloh. K jednotlivým úlohám je vhodné si robiť stručný záznam (krátku dokumentáciu) pre ľahšie odovzdávanie projektu. Dáta k úlohám sú uložené v AISE.

Úloha 5a:

Vytvorte MLP sieť na rozpoznanie (klasifikáciu) objektov, konkrétne 5 skupín bodov (každá skupina 50 bodov) na základe troch parametrov x, y, z . Dáta sú uložené v súbore *databody.mat*.

Na vytvorenie MLP siete na klasifikáciu použite funkciu *patternet* – (vid'. ukážka *demo_nn2mlp.m*). Na riešenie úlohy máte pripravený program *uloha5a.m*, ktorý treba doplniť podľa zadania úlohy. Na tréningovanie použite náhodný výber dát, maximálne 80% z celkového počtu dát, zvyšné dáta použite na testovanie a validáciu (alebo len testovanie).

Nájdite čo najmenší počet neurónov v skrytej vrstve a ukončovacie podmienky tréningovania, tak aby NS správne klasifikovala čo najviac bodov (najlepšie všetky vzorky bodov, akceptované sú 2 zle klasifikované). (vid'. kontingenčná matica – confusion matrix *plotconfusion*). Zobrazte priebeh chyby siete v procese učenia.

Na záver otestujte 5 definovaných bodov v programe a zatried'te ich do príslušných skupín pomocou NS. Vyčísľte výstup NS a zatried'te def. body do skupín. Príslušnosť do skupín aj graficky znázornite.

Úloha 5b:

Vytvorte NS typu viacvrstvová perceptrónová sieť – MLP sieť (*fitnet* – vid'. ukážka *demo_nn3fitnet.m*), ktorá umožní aproximáciu nelineárnej funkcie f s jedným vstupom x a výstupom y , t.j. $y=f(x)$. Na riešenie úlohy máte pripravený program *uloha5b.m*, ktorý treba doplniť podľa zadania úlohy.

Na tréningovanie NS použite merané body uložené v súbore *datafun.mat*, v ktorom sa nachádzajú nasledovné premenné:

% x, y – vstupná a výstupná premenná funkcie pre tréningovanie siete, zobrazenie (plot)

% $indx_train, indx_test$ – indexy pre indexové rozdelenie dát na tréningové a testovacie dáta

Nájdite optimálnu štruktúru MLP siete (najmenší počet neurónov v skrytej vrstve) a nastavenie parametrov tréningovania, tak aby ste dosiahli minimálnu odchýlku na testovacích dátach (pod $1e-4$) a tvar funkcie vypočítaný z NS sa zhodoval s nameranými dátami. Zdokumentujte postup tréningovania NS. (Zobraziť priebeh chyby na tréningových aj testovacích dátach)

Ukážte štruktúru NS a proces tréovania, (malý a veľký počet neurónov v skrytej vrstve), kde sa sieť dobre nenaučila. Zdôvodnite prečo.

Nastavte indexové rozdelenie dát (viď. index premenné v datovom súbore) a zrealizujte tréovanie siete.

Vyčísľte chyby SSE (suma kvadrátov odchýliek medzi meraným výstupom a výstupom siete), MSE (priemer z SSE), MAE (maximálna absolútna odchýlka medzi meraným výstupom a výstupom siete) na tréovacích a testovacích dátach.

Graficky porovnajte pôvodné data s výstupom NS, v grafe farebne označte tréovacie, testovacie dáta a výstup neurónovej siete.

Úloha 6 :

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná klasifikovať (rozpoznávať) stav bábätka na základe CTG vyšetrenia (tzv. Kardiotokografia). CTG je vyšetrenie, ktoré **zaznamenáva ozvy srdca bábätka**. CTG je predovšetkým neinvazívne a najvyšš bezpečné vyšetrenie, skúsený pôrodník z neho vie vyčítať napríklad aj to, či vám nehrozí pôrod pred termínom, a aký je celkový stav bábätka. Okrem toho ukazuje, či má dieťa dostatok kyslíka. Výsledkom vyšetrenia býva kardiozáznam, ktorý ukazuje ozvy srdiečka a zároveň sťahy maternice. Z kardiozáznamu boli vypočítané určité vybrané parametre (príznaky), na základe ktorých boli kardiozáznamy zatriedené do 10 tried. Záznamy sú tiež zatriedené do troch tried (1-normálny, 2-podozrivý, 3-patologický). Pre klasifikáciu boli vytvorené dáta z 2126 meraní, z ktorým bolo vypočítaných 25 parametrov z meraných signálov. Všetky data sa nachádzajú v súbore **CTGdata.mat**, kde vstupné parametre ochorenia sú v premennej **NDATA** a skupiny ochorenia sa nachádzajú v premennej **typ_ochorenia** (**1-normálny, 2-podozrivý, 3-patologický**).

Úlohou siete je zatriediť kardiozáznam z CTG do 3 tried. Na týchto dátach vykonajte tréovanie siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné data použite na testovanie. Štruktúru siete ako aj parametre tréovania nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie na testovacích dátach. Pre natréňovanú NS vyčísľte percentuálnu úspešnosť klasifikácie (**plotconfusion**) na tréovacích, testovacích a celkových dátach. Priemerná percentuálna úspešnosť klasifikácie na testovacích dátach musí byť väčšia ako 92%. V úlohe zdokumentujte štruktúru siete, proces tréovania NS a dosiahnuté úspešnosti pre **najlepšiu sieť**. Vykonajte testovanie NS metódou krížovej validácie, alebo 5 krát spustíte tréovanie pri náhodnom rozdelení dát. Vyčísľte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréovacích a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Pre najlepšie natréňovanú sieť otestujte z každého typu ochorenia aspoň jednu vzorku. Vyčísľte výstup NS a zatriedte ochorenia do skupín.

Úloha 7:

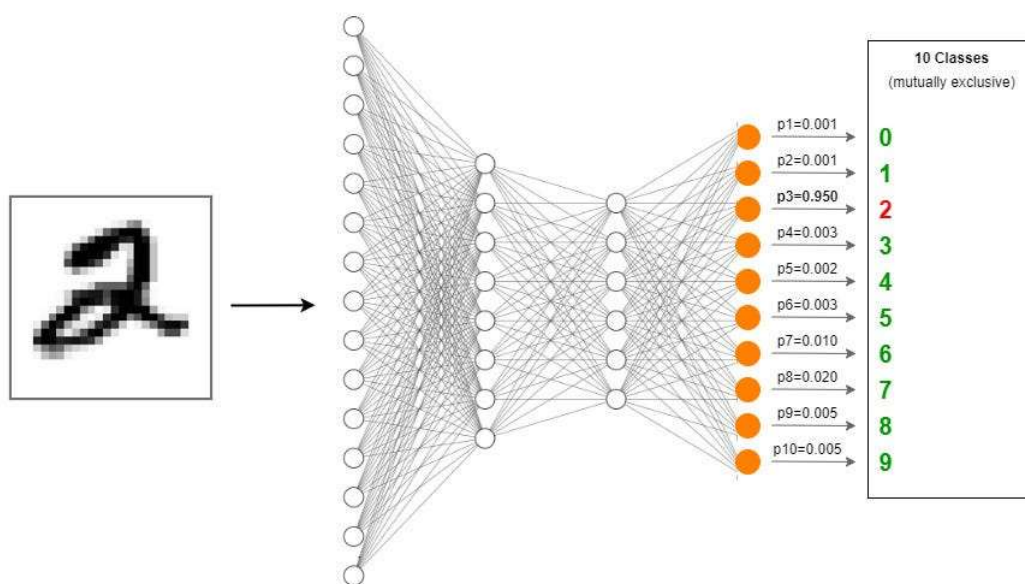
Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná rozpoznávať rukou písané číslice. Na tréovanie MLP siete použite dataset znakov MNIST (obr.1) (<http://yann.lecun.com/exdb/mnist/>).



Obr. 1 Dataset rukou písaných číslíc MNIST

Písmená abecedy sú nasnímané v rastrí 28x28, t.j. NS bude mať 784 vstupov s rozsahom $\langle 0;1 \rangle$, kde hodnota 0 predstavuje biele a 1 čierne políčko. Údaje ku písmenám sa nachádzajú v súbore ***datapiscisla_all.mat*** v premennej ***XDataall***.

```
>> load datapiscisla_all           % načítanie dát
>> dispznak(XDataall(:,1),28,28); % vykreslenie 1. znaku
```



Obr. 2 Principiálna schéma rozpoznávania rukou písaných číslíc pomocou MLP siete

Pri vytváraní a tréňovaní MLP siete môžete použiť maximálne 60% z celkového počtu dát, zvyšné data použité na testovanie a validáciu. Pre natréňovanú NS vyčísľte percentuálnu úspešnosť klasifikácie. (***plotconfusion***). Nájdite vhodnú štruktúru siete a parametre tréňovania. Celková percentuálna úspešnosť klasifikácie musí byť väčšia ako 95%. V úlohe zdokumentujte štruktúru siete, proces tréňovania NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo aspoň 5 krát spustíte tréňovanie pri náhodnom rozdeľovaní dát. Vyčísľte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréňovacích a testovacích dátach. Otestujte z každého čísla aspoň jednu vzorku. Vyčísľte výstup NS a zatried'te čísla do skupín.

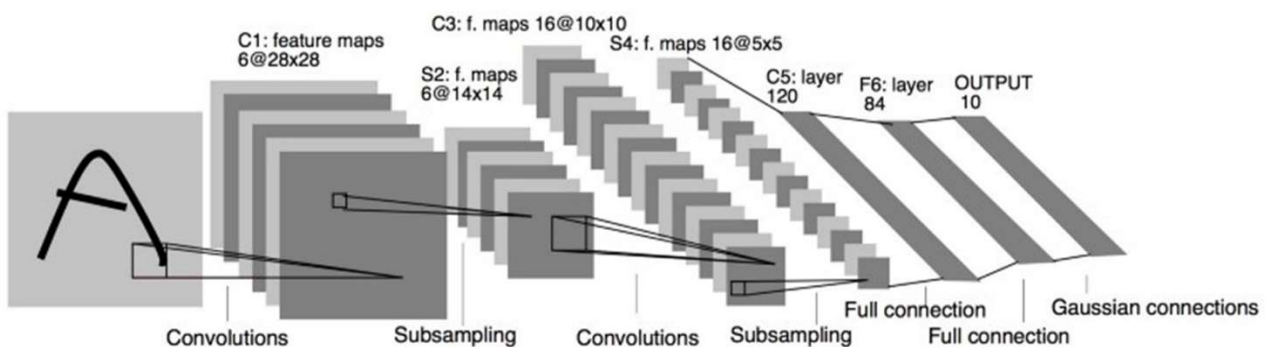
Úloha 8:

Vytvorte a natrénujte konvolučnú neurónovú sieť (CNN), ktorá bude schopná rozpoznávať rukou písané číslice, tak isto ako v úlohe 7 s MLP sieťou. Postupujte podľa demonštračnej ukážky v Matlabe „Create simple image classification network“.

Pri vytváraní a trénovaní CNN siete použite 60% z celkového počtu dát, zvyšné data použite na testovanie. Pre natrénovanú NS vyčísľte percentuálnu úspešnosť klasifikácie. (**confusion**, **plotconfusion**). (Zvlášť na trénovacích a testovacích dátach). Nastavte štruktúru siete a parametre trénovania, tak aby ste dosiahli čo najväčšiu percentuálnu úspešnosť klasifikácie na testovacích dátach. Pre naladenú štruktúru siete vykonajte 5 spustení s náhodným rozdelením dát alebo krížovú validáciu. Pre najlepšie natrénovanú CNN zobrazte aj kontingenčnú maticu.

Vykonajte porovnanie priebehov trénovania, chyby a úspešnosti dvoch vybraných štruktúr CNN siete na trénovacích a testovacích dátach pre rozdelenie 60%/40%.

Vykonajte porovnanie priebehov trénovania, chyby a úspešnosti MLP a CNN siete na trénovacích a testovacích dátach pre rozdelenie 60%/40%.



Obr. 3 Principiálna schéma rozpoznávania rukou písaných číslic pomocou CNN siete

Čo treba vedieť k jednotlivým úlohám z projektu NS.

Úloha 5a

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na trénovacie, testovacie alebo aj validačné data.
- Popis štruktúry MLP siete (vstupy, výstupy, typy neurónov v jednotlivých vrstvách, ich počet).
- Nastavené parametre trénovania (ukončovacie podmienky, použitá kritériálna funkcia)
- Stačí jedna čo najjednoduchšia štruktúra NS, čo spĺňala zadanie úlohy.
- Graf priebehu procesu trénovania aj s komentárom
- Kontingenčná matica (**plotconfusion**) aj s komentárom
- Postup testovania vybraných 5 bodov.

Úloha 5b

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na tréningové, testovacie data. Upozornenie máte použiť indexové rozdelenie.
- Popis štruktúry MLP siete (vstupy, výstupy, typy neurónov v jednotlivých vrstvách, ich počet).
- Nastavené parametre tréningovania (ukončovacie podmienky, použitá kritériálna funkcia)
- Najprv nastaviť počet neurónov v skrytej vrstve NS, čo spĺňa kvalitu aproximácie funkcie. Potom dať podstatne menší a väčší počet neurónov a zdokumentovať výsledky
- Grafy priebehu procesu tréningovania aj s komentárom pre 3 nastavenia neurónov
- Porovnajte priebehy výstupov funkcie a NS a vyčíslite chyby SSE, MSE, MAE pre 3 nastavenia neurónov.

Úloha 6

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na tréningové, testovacie alebo aj validačné data.
- Popis štruktúry MLP siete (vstupy, výstupy, typy neurónov v jednotlivých vrstvách, ich počet).
- Nastavené parametre tréningovania (ukončovacie podmienky, použitá kritériálna funkcia)
- Najprv nastaviť počet neurónov v skrytej vrstve NS, čo spĺňa požiadavky klasifikácie. Zdokumentovať výsledky pre najlepšie tréningovanie. (proces tréningovania, kontingenčná matica)
- Pri tomto nastavení parametrov urobiť krížovú validáciu alebo 5 krát spustiť tréningovanie s náhodným rozdelením dát a vyčíslit' min, max, priemernú úspešnosť klasifikácie.
- Potom ukázať dve iné nastavenie parametrov NS a zdokumentovať výsledky (proces tréningovania, kontingenčná matica) (stačí jedno tréningovanie)
- Postup testovania vybraných vzoriek.
- Pre najlepšie natrénovanú sieť, vypočítať úspešnosť klasifikácie, senzitivitu a špecificitu. (1. skupina – negatívny, 2. a 3. skupina – pozitívny)

Úloha 7

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na tréningové, testovacie alebo aj validačné data.
- Popis štruktúry MLP siete (vstupy, výstupy, typy neurónov v jednotlivých vrstvách, ich počet).

- Nastavené parametre tréovania (ukončovacie podmienky, použitá kritériálna funkcia)
- Najprv nastaviť počet neurónov v skrytej vrstve NS, čo spĺňa požiadavky klasifikácie. Následne ladiť parametre tréovania. Zdokumentovať výsledky pre najlepšie tréovanie. (proces tréovania, kontingenčná matica)
- Pri tomto nastavení parametrov urobiť krížovú validáciu alebo 5 krát spustiť tréovanie s náhodným rozdelením dát a vyčísliť min, max, priemernú úspešnosť klasifikácie.
- Vedieť ukázať aj iné nastavenie parametrov NS
- Postup testovania vybraných vzoriek.

Úloha 8

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na tréovacie, testovacie alebo aj validačné data.
- Popis štruktúry CNN siete (vstupy, výstupy, jednotlivé vrstvy).
- Nastavené parametre tréovania (ukončovacie podmienky, použitá kritériálna funkcia, krok učenia, veľkosť dávky, ...)
- Najprv zvoliť architektúru CNN (2 alebo 3 konvolučné vrstvy), čo spĺňa požiadavky klasifikácie. Následne ladiť parametre tréovania. Zdokumentovať výsledky pre najlepšie tréovanie. (proces tréovania, kontingenčná matica, úspešnosť klasifikácie tréovacie a testovacie data)
- Pri tomto nastavení parametrov minimálne 3 krát (ideálne 5 krát) spustiť tréovanie s náhodným rozdelením dát a vyčísliť min, max, priemernú úspešnosť klasifikácie.
- Vedieť ukázať aj iné nastavenie parametrov CNN
- Porovnať dve štruktúry CNN. (proces tréovania, kontingenčná matica, úspešnosť klasifikácie tréovacie a testovacie data)
- Porovnať MLP a CNN sieť. (proces tréovania, kontingenčná matica, úspešnosť klasifikácie tréovacie a testovacie data)
- Postup testovania vybraných vzoriek.