

Zadanie projektu č.2 – Neurónové siete

Vyriešte jednu z určených úloh podľa inštrukcií učiteľa. Výsledky riešenia prezentujte pred celou skupinou. Súčasťou prezentácie je stručná charakteristika problému, popis štruktúry neurónovej siete, naznačenie postupu tréningu siete, parametre tréningu, graf priebehu učenia siete, kontingenčná tabuľka úspešnosti klasifikácie (rozpoznávania).

1. Úloha

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná rozpoznávať rukou písané číslice.

Na tréning MLP siete použite dataset znakov MNIST (obr.1)

(<http://yann.lecun.com/exdb/mnist/>).



Obr. 1 Dataset rukou písaných číslic MNIST

Písmená abecedy sú nasnímané v rastri 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
```

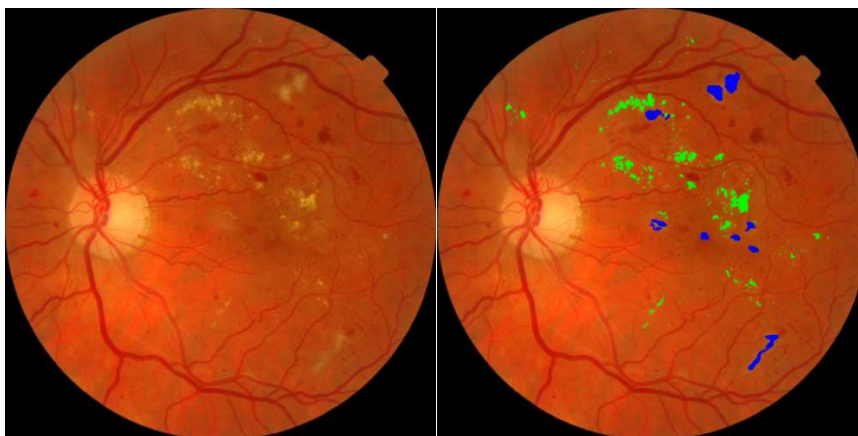
Pri vytváraní a tréňovaní MLP siete môžete použiť maximálne 60% z celkového počtu dát, zvyšné data použite na testovanie a validáciu. Pre natréňovanú NS vyčísľte percentuálnu úspešnosť klasifikácie. (***plotconfusion***). Nájdite vhodnú štruktúru siete a parametre tréningu. Celková percentuálna úspešnosť klasifikácie musí byť väčšia ako 95%. V úlohe zdokumentujte štruktúru siete, proces tréningu NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo aspoň 10 krát spustíte tréning pri náhodnom rozdeľovaní dát. Vyčísľte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréningových a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Otestujte z každého čísla aspoň jednu vzorku (bližšie informácie cvičiaci).

2. Úloha

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná klasifikovať (rozpoznávať) ochorenie očného pozadia Diabetickú retinopatiu, vyskytujúcu sa pri pacientoch s ochorením na cukrovku. Na obrázku môžete vidieť príklad poškodenia očného pozadia. Vytvorené dáta na klasifikáciu, rozpoznanie ochorenia diabetickej retinopatie, boli získané z databázy obrázkov MESSIDOR. Pre klasifikáciu bolo z obrázkov získaných 19 parametrov na základe, ktorých sú obrázky klasifikované do dvoch skupín. (0 – žiadne ochorenie, 1 – diagnostikované ochorenie diabetickej retinopatie 1.,2. alebo 3 stupňa). Popis jednotlivých vstupov nájdete:

<https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set>

Na týchto dátach vykonajte tréning siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné dáta použijete na testovanie a validáciu. Štruktúru siete ako aj parametre tréningu nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie. Pre natréňovanú NS vyčísľte percentuálnu úspešnosť klasifikácie (*plotconfusion*). Celková percentuálna úspešnosť klasifikácie musí byť väčšia ako 80%. V úlohe zdokumentujte štruktúru siete, proces tréningu NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo aspoň 10 krát spustíte tréning pri náhodnom rozdeľovaní dát. Vyčísľte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréningových a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Otestujte náhodne vybraných 10 vzoriek. (bližšie informácie cvičiaci).



Obr. 2 Originálny a obrázok označený oftalmológom

3. Úloha

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná klasifikovať (hodnotiť) kvalitu vína. Klasifikácia, určenie kvality vína je realizované na základe 11 nameraných chemických parametrov vína (napr. kyslosť, pH, zvyškový cukor, chloridy, oxid syričitý, hustota, sulfáty, alkohol). Bližšie informácie o dátach:

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Na týchto dátach vykonajte tréning siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné dáta použijete na testovanie a validáciu. Štruktúru siete ako aj parametre tréningu nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie. Pre natréňovanú

NS vyčíslíte percentuálnu úspešnosť klasifikácie (*plotconfusion*). Celková percentuálnu úspešnosť klasifikácie musí byť väčšia ako 84%. V úlohe zdokumentujte štruktúru siete, proces tréovania NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo aspoň 10 krát spustíte tréovanie pri náhodnom rozdeľovaní dát. Vyčíslíte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréovacích a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Otestujte z každej skupiny aspoň jednu vzorku (bližšie informácie cvičiaci).

4. Úloha

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná rozpoznávať veľké písmená od A po Z (znaky) abecedy. Klasifikácia znakov je realizované na základe 16 nameraných parametrov znaku (napr. šírka, výška, počet pixelov, priemerný počet pixelov v smere x a y, ...). Bližšie informácie o dátach:

<https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>

Na týchto dátach vykonajte tréovanie siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné data použite na testovanie a validáciu. Štruktúru siete ako aj parametre tréovania nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie. Pre natréovanú NS vyčíslíte percentuálnu úspešnosť klasifikácie (*plotconfusion*). Celková percentuálnu úspešnosť klasifikácie musí byť väčšia ako 95%. V úlohe zdokumentujte štruktúru siete, proces tréovania NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo aspoň 10 krát spustíte tréovanie pri náhodnom rozdeľovaní dát. Vyčíslíte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréovacích a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Otestujte z každého znaku aspoň jednu vzorku (bližšie informácie cvičiaci).

5. Úloha

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná klasifikovať (rozpoznávať) ochorenie srdca, konkrétne arytmie, ktorá bola zatriedená do 10 skupín. Pre klasifikáciu boli vytvorené dáta od 442 pacientov u ktorým bolo zaznamenaných 227 parametrov (osobné ako vek, pohlavie, ... a vyhodnotenie z EKG signálov). Na základe týchto 227 parametrov bol pacientom diagnostikovaný typ ochorenia srdca, ktoré boli zatriedené do nasledovných do 10 skupín. (1 – žiadne ochorenie, 2 – ischemická choroba, 3 – starý infarkt myokardu, ..., 9 – spojená skupina ochorení, 10 – ostatné ochorenia srdca). Popis jednotlivých vstupov a výstupov je v súbore *arrhythmia_description.txt*. Všetky data sa nachádzajú v súbore *dataarytmiasrdca.mat*, kde vstupné parametre ochorenia sú v premennej *NDATA* a skupina ochorenia sa nachádza v premennej *typ_ochorenia*.

Na týchto dátach vykonajte tréovanie siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné data použite na testovanie a validáciu. Štruktúru siete ako aj parametre tréovania nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie. Pre natréovanú NS vyčíslíte percentuálnu úspešnosť klasifikácie (*plotconfusion*). Celková percentuálnu úspešnosť klasifikácie musí byť väčšia ako 75%. V úlohe zdokumentujte štruktúru siete, proces tréovania NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou

krížovej validácie, alebo aspoň 10 krát spustite tréovanie pri náhodnom rozdelovaní dát. Vyčísľte minimálnu, maximálnu a priemernú úspešnosť klasifikácie na tréovacích a testovacích dátach. Porovnajte aspoň 3 rôzne štruktúry siete (rôzne počty neurónov). Otestujte z každého typu ochorenia aspoň jednu vzorku (bližšie informácie cvičiaci).