

Zadanie projektu – Umelé neurónové siete

Úlohy:

4. Klasifikácia do skupín pomocou viacvrstvovej perceptronovej (MLP) siete. (5 b)

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

6. Praktická úloha rozpoznávania pomocou viacvrstvovej perceptronovej (MLP) siete. (8 b)

Úlohy 4 a 5 predstavujú jednoduchšie príklady na aproximáciu a klasifikáciu pomocou viacvrstvovej perceptronovej siete. Vašou úlohou je modifikovať demo ukážky k prednáške 6 (2. prednáška NS) podľa zadaných úloh. Výsledky úloh podľa zadania z dokumentovať do krátkej dokumentácie projektu. Dáta k úlohám sú uložené v AISe.

Úloha 4 :

Vytvorte MLP sieť na rozpoznanie (klasifikáciu) objektov, konkrétne do 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 trénovanie 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 trénovania, 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 5 :

Vytvorte NS typu viacvrstvová perceptronová 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 trénovanie NS použite merané body uložené v súbore *datafun.mat*, v ktorom sa nachádzajú nasledovné premenné:

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

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

Nájdite optimálnu štruktúru MLP siete (najmenší počet neurónov v skrytej vrstve) a nastavenie parametrov trénovania, 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 trénovania NS. (Zobraziť priebeh chyby na trénovacích aj testovacích dátach)

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

Nastavte indexové rozdelenie dát (vid'. index premenné v datovom súbore) a zrealizujte trénovanie 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 tréningových a testovacích dátach.

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

Úloha 6 :

Vytvorte a natrénujte neurónovú sieť, ktorá bude schopná klasifikovať (rozpoznávať) typy ochorení identifikované na základe CTG vyšetrenia (tzv. Kardiotokografia). CTG je vyšetrenie, ktoré **zaznamenáva ozvy srdca bábätká**. 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 bábätká. 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 dáta sa nachádzajú v súbore **CTGdata.mat**, kde vstupné parametre ochorenia sú v premennej **NDA** 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 tréningovanie siete, môžete použiť maximálne 60% z celkového počtu dát, zvyšné dáta použite na testovanie. Štruktúru siete ako aj parametre tréningovania nastavte tak, aby ste dosiahli, čo najlepšiu percentuálnu úspešnosť klasifikácie na testovacích dátach. Pre natréňovanú NS vyčísľte percentuálnu úspešnosť klasifikácie (**plotconfusion**) na tréningových, testovacích a celkových dátach. Priemerná percentuálna úspešnosť klasifikácie na testovacích dátach musí byť väčšia ako 92%. V úlohe zdokumentujte štruktúru siete, proces tréningovania NS a dosiahnuté úspešnosti pre najlepšiu sieť. Vykonajte testovanie NS metódou krížovej validácie, alebo 10 krát spustíte tréningovanie 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). Pre najlepšie natréňovanú sieť otestujte z každého typu ochorenia aspoň jednu vzorku. Vyčísľte výstup NS a zatriedte ochorenia do skupín.

Čo treba vedieť k jednotlivým úlohám z projektu NS.

4. Úloha

- 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 spĺňa zadanie úlohy.
- Graf priebehu procesu trénovania aj s komentárom
- Kontingenčná matica (*plotconfusion*) aj s komentárom
- Postup testovania vybraných 5 bodov.

5. Úloha

- Stručne vlastnými slovami, čo bolo úlohou (max. 3 vety).
- Popis vstupných a výstupných dát, ako ste ich rozdelili na trénovacie, testovacie data. Upozornenie máte použiť indexové rozdelenie.
- Popis štruktúry MLP siete (vstupy, výstupy, typy neurónov v jednotlivých vrstvách, ich počet).
- Nastavené parametre trénovania (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 trénovania aj s komentárom pre 3 nastavenia neurónov
- Porovnajte priebehy výstupov funkcie a NS a vyčíslite chyby SSE, MSE, MAE pre 3 nastavenia neurónov.

6. Úloha

- **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)**
- **Najprv nastaviť počet neurónov v skrytej vrstve NS, čo spĺňa požiadavky klasifikácie. Zdokumentovať výsledky pre najlepšie trénovanie. (proces trénovania, kontingenčná matica)**
- **Pri tomto nastavení parametrov urobiť krížovú validáciu alebo 10 krát spustiť trénovanie s náhodným rozdelením dát a vyčíslit' min, max, priemernú úspešnosť klasifikácie.**
- **Potom ukázať dve iné nastavenie parametrov NS a zdokumentovať výsledky (proces trénovania, kontingenčná matica) (stačí jedno trénovanie)**
- **Postup testovania vybraných vzoriek.**
- **Pre najlepšie natrénovanú sieť, vypočítať úspešnosť klasifikácie, senzitivitu a špecificitu. (1. skupina – negatívny, 2. a 3. skupina – pozitívny)**