

Y33AUI: Lineární diskriminační funkce I.

Petr Pošík

Katedra kybernetiky

ČVUT FEL

1 Perceptronový algoritmus

1. Zopakujte si poznatky o perceptronovém algoritmu z přednášek. Jak pracuje?
2. Naimplementujte perceptronový algoritmus v MATLABu.
3. Otestujte jej na reálných datech (digitalizované ručně psané číslice).

2 Pomocné funkce

1. Vytvořte funkci `homog`, která z nehomogenních souřadnic bodů udělá souřadnice homogenní, tj. přidá ke každému bodu $D + 1$. souřadnici s hodnotou 1. Jako vstup tedy bude mít
 - matici $[D \times N]$, tedy sadu N bodů, každý o D souřadnicích,a jako výstup bude produkovat
 - matici $[(D + 1) \times N]$, tedy sadu N bodů, každý o $D + 1$ souřadnicích.
2. Vytvořte funkci `mm`, která bude vytvářet tzv. *misclassification matrix*, klasifikační matici, který se používá pro posouzení kvality modelu u klasifikace. Objekty mohou pocházet z různých tříd, klasifikátor je může zařadit buď do správné třídy, nebo do některé z nesprávných tříd. Klasifikační matice udává, kolik objektů ze třídy 1 zařadil klasifikátor do třídy 1, do třídy 2, ..., kolik objektů ze třídy 2 zařadil klasifikátor do třídy 1, do třídy 2, ... Funkce bude mít dva vstupy:
 - skutečné zařazení objektů do tříd (vektor $[1 \times N]$) a
 - zařazení do tříd predikované klasifikátorem (opět vektor $[1 \times N]$).

Výstupem bude

- matice $M [C \times C]$, kde C je počet tříd, taková, že $M(1,1)$ bude počet objektů třídy jedna správně zařazených do třídy 1, $M(1,2)$ bude počet objektů třídy 2 nesprávně zařazených do třídy 1, atd.

3 Lineární diskriminační funkce a její učení

1. Vytvořte funkci `predLinDiscrFun`, která jako parametry převezme

- lineární model, tedy vektor vah $w \in [D+1, 1]$ a
- neohodnocené body (matice $X \in [D, N]$)

a jako výsledek na výstupu poskytně

- predikce třídy, pro všechny body (vektor $y \in [1, N]$).

Domluvme se, že třídy budeme označovat kódy "1" a "-1".

2. Vytvořte funkci `fitLinDiscrFunUsingPerceptron` pro učení lineární diskriminační funkce pomocí perceptronového algoritmu, která jako parametry převezme trénovací množinu bodů, tj.

- souřadnice bodů (matice $X \in [D \times N]$) a
- vektor tříd, do níž jednotlivé body patří (vektor $y \in [1, N]$)

a jako výsledek vrátí

- vektor parametrů lineární diskriminační funkce (vektor vah $w \in [D+1, 1]$)

4 Testování perceptronového algoritmu

Otestujte perceptronový algoritmus na úloze rozpoznávání digitalizovaných ručně psaných čísel:

1. zdroj dat: UCI Data Repository, datové soubory si stáhněte, pro načtení do MATLABu můžete použít např. funkci `dlmread`
2. každé číslo je popsáno maticí 8×8 bodů v odstínech šedi
3. definujte si nějakým způsobem dvě třídy číslic a pokuste se vytvořit klasifikátor, který tyto třídy odděluje:
 - omezte se např. pouze na číslice 1 a 8 a pokuste se je oddělit lineárním klasifikátorem
 - použijte všechny číslice a zkuste oddělit číslice 0–4 od číslic 5–9

5 Úkoly

1. Pokuste se vytvořit klasifikátor metodou nejbližších sousedů a aplikujte jej na stejné úlohy, jako lineární klasifikátor.
 - Vytvořte funkci `predNN`, která bude predikovat třídu na základě modelu a souřadnic nových bodů. Jak zde vypadá model? Co je u metody nejbližších sousedů modelem? Co je třeba také specifikovat, abychom mohli definovat, co znamená "nejbližší"?
2. Oba typy modelů porovnejte.