

Úloha 11 – Lineární klasifikátor

Zadání

- Implementujte perceptronový algoritmus podle slajdů z přednášek.
- Pomocí implementovaného algoritmu klasifikujte OBA datasety ze stránek předmětu.
- Proveďte rozšíření báze (pro $s = 2$) a opět klasifikujte oba datasety.

Obsah zprávy

- stručný popis algoritmu,
- klasifikační úspěšnost lineárního klasifikátoru a lineárního klasifikátoru s rozšířenou bází pro oba datasety,
- zobrazení rozhodovací hranice pro vybrané případy,
- váš komentář ke všem předchozím bodům.

Algoritmus

Perceptronový algoritmus implementuji funkcí `[weights, resTraindata] = perceptron(data, class, rozsireni, prah)`. Vstupním parametrem je množina trénovacích dat, množina trénovacích tříd, hodnota false/true, která reprezentuje rozšíření báze a posledním parametrem je hodnota pro nastavení prahu perceptronu.

Funkce vrací nastavení `weights`, které použijeme pro klasifikaci testovacích dat a pro zobrazení dělicí čáry v datech.

Nejprve se provede inicializační fáze, kdy se inicializují potřebné proměnné. Pro limit iterací algoritmu jsem zvolila hodnotu 500.000 a hodnotu proměnné `stejneWeights` jsem nastavila na 100. Tzn., pokud 100x za sebou bude stejná hodnota `weights`, algoritmus bude ukončen.

Pokud je proměnná `rozsireni` nastavena na `false`, inicializace `weights` je pro 2 dimenze a původní báze zůstává beze změny. Pokud je proměnná `rozsireni` nastavena na `true`, dochází k rozšíření báze. Inicializace `weights` je pro 5 dimenzí a pro rozšíření báze je zavolána funkce `expandData`.

Lineární klasifikátor bez rozšíření báze

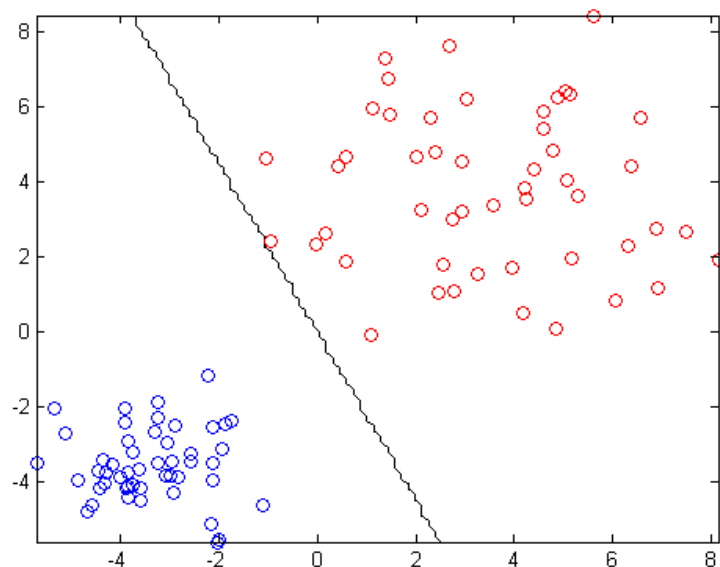
Výsledky pro lineárně separovatelná data `linsep.mat`:

Test na trénovací množině dat:

```
[weights, resTraindata] = perceptron(traindata, trainclass, false, 0);
```

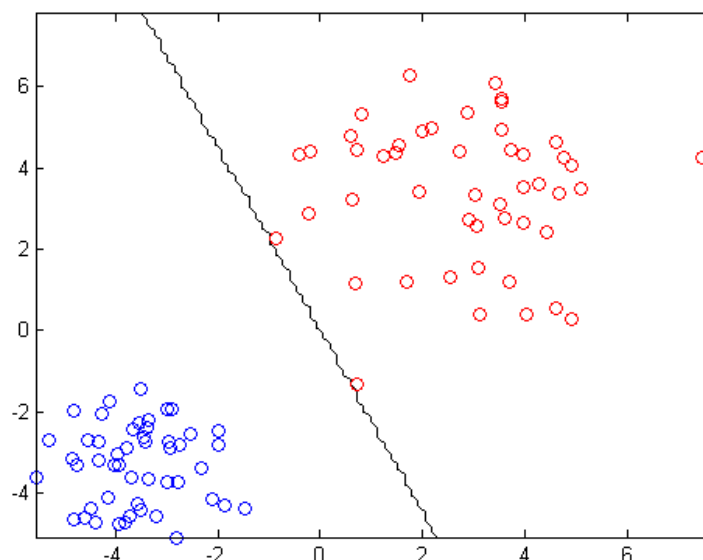
Algoritmus skončil po 121 iteracích.

```
linearplot(resTraindata, trainclass, weights);
```



Z obrázku patrné, že data jsou lineárně separovatelná, protože došlo k jejich rozdělení na dvě skupiny.

Klasifikace testování množiny dat - `linearplot(testdata, testclass, weights);`



I zde je uspokojivá úspěšnost, protože opět došlo k úspěšnému rozdělení dat na dvě části.

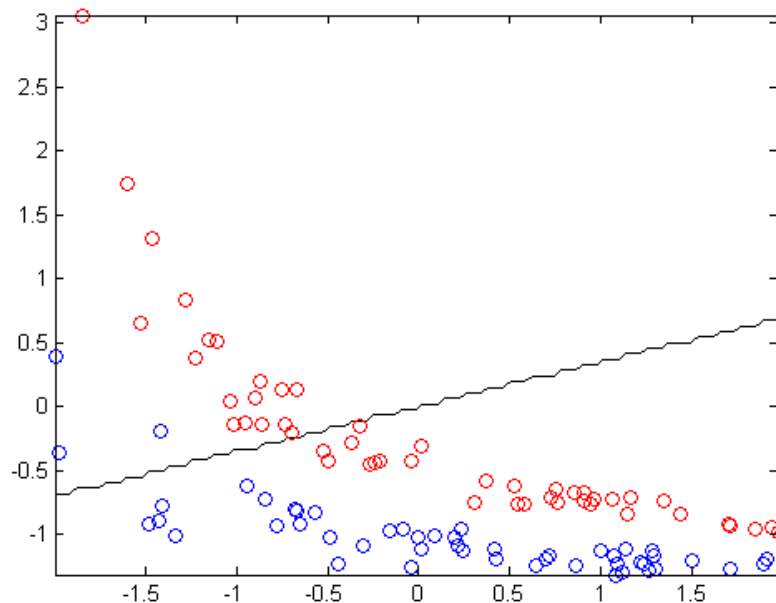
Výsledky pro lineárně neseparovatelná data `nonlinsep.mat`:

Test na trénovací množině dat:

```
[weights, resTraindata] = perceptron(traindata, trainclass, false, 0);
```

Test skončil neúspěchem, protože ani po stanovených 500 000 iterací se nepodařilo data separovat. Neseparovatelnost dat je patrná i z následujícího obrázku.

`linearplot(resTraindata, trainclass, weights);`



Vzhledem k tomu, že pro trénovací data skončil test neúspěšně, je bezpředmětné klasifikovat také množinu testovací data.

Lineární klasifikátor s rozšířením báze

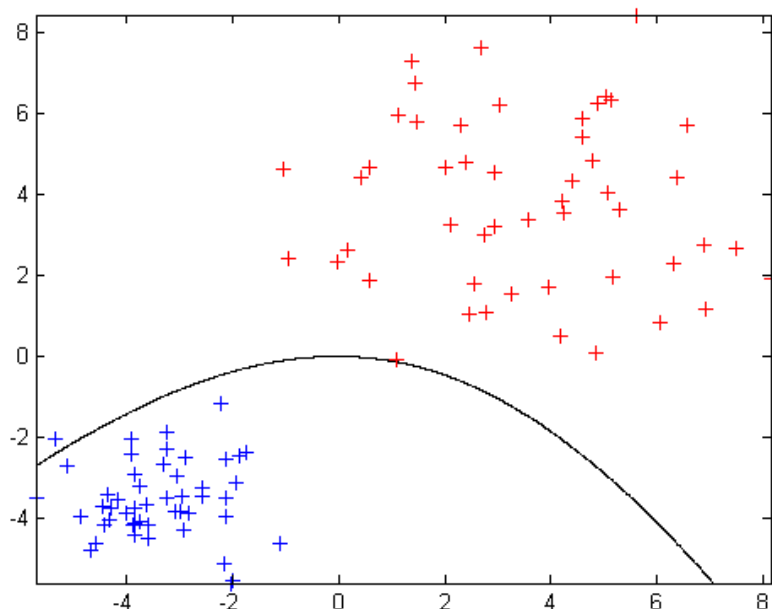
Výsledky pro lineárně separovatelná data `linsep.mat`:

Test na trénovací množině dat:

```
[weights, resTraindata] = perceptron(traindata, trainclass, true, 0);
```

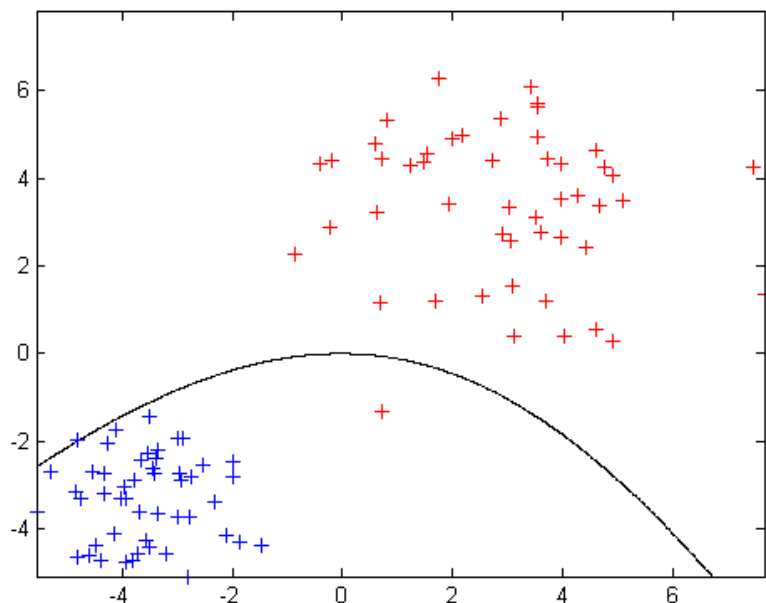
Algoritmus skončil po 144 iteracích.

```
expandedPlot(resTraindata, trainclass, weights, 0);
```



Množinu dat se podařilo úspěšně separovat. K deformaci dělící čáry došlo v důsledku rozšíření dimenze ze 2 na 5.

Klasifikace testovacích dat - `expandedPlot(expandData(testdata), testclass, weights, 0);`



U množiny testovacích dat byla data rozdělena s jednou chybou – jeden červený bod byl zařazen do skupiny modrých. I přesto lze ale výsledek hodnotit jako uspokojivý.

Výsledky pro lineárně neseparovatelná data **nonlinsep.mat**:

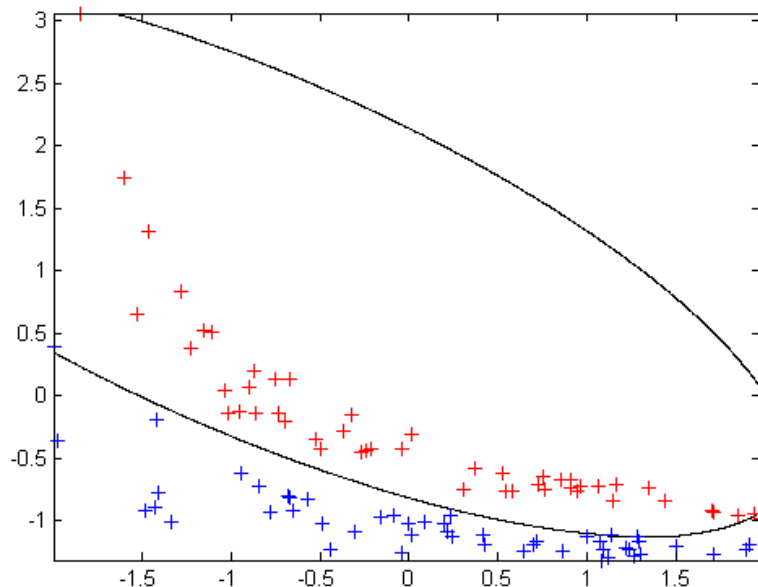
Do parametrů funkce je nutno nastavit práh, aby bylo možno data separovat.
Zvolila jsem libovolnou hodnotu 150.

Test na trénovací množině dat:

```
[ weights, resTraindata ] = perceptron( traindata, trainclass, true, 150);
```

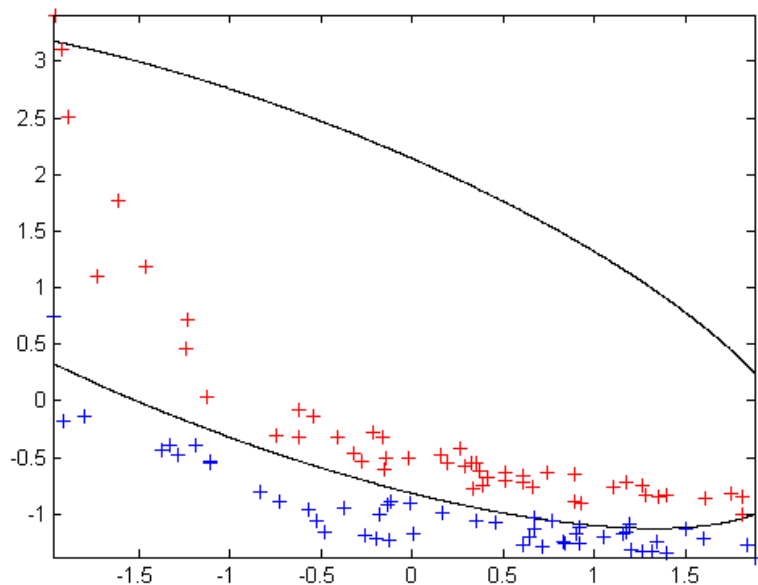
Algoritmus skončil po 144 iteracích.

```
expandedPlot(resTraindata, trainclass, weights, 150);
```



Z obrázku je patrné, že data byla úspěšně rozdělena i po rozšíření báze.

Klasifikace testovacích dat - `expandedPlot(expandData(testdata), testclass, weights, 150);`



I pro testovací množinu dat dopadl test úspěšně. Všechna data byla úspěšně rozdělena.

Závěr

Pomocí implementace perceptronového algoritmu jsem oklasifikovala oba zadané datasety. Ověřila jsme si, že dataset **nonlinsep.mat** je lineárně neseparovatelný. Pokud jsem ale pomocí transformace rozšířila bázi a zkusila jsme data separovat pomocí křivky, byla množina dat v 5-ti dimenzionálním prostoru úspěšně rozdělena.