

Vytěžování dat, cvičení 11: Lineární klasifikátor

Miroslav Čepek, Michael Anděl



Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta elektrotechnická, ČVUT

Perceptronový algoritmus

Let \mathbf{x}_i be the i -th sample from the feature space and y_i its class indicator, then finding a linear classifier with *zero training error* is solving:

$$\mathbf{w}\mathbf{x}_i + b \geq 0, \forall i : y_i = 1,$$

$$\mathbf{w}\mathbf{x}_i + b < 0, \forall i : y_i = 2,$$

After a simple data transformation:

$$\mathbf{v} \leftarrow [\mathbf{w}, b],$$

$$\mathbf{x}_i^* \leftarrow [\mathbf{x}_i, 1], \forall i : y_i = 1,$$

$$\mathbf{x}_i^* \leftarrow -[\mathbf{x}_i, 1], \forall i : y_i = 2,$$

we can solve the problem with respect to one condition only, i.e., to $\mathbf{v}\mathbf{x}_i^* > 0, \forall i = 1 \dots N$, using following algorithm:

Data: $\{\mathbf{x}_i^*\}_{i=1}^N$ transformed data samples

Result: \mathbf{v} weights of the classifier

begin

$\mathbf{v} \leftarrow 0;$

while $\exists \mathbf{x}_i : \mathbf{v}\mathbf{x}_i^* \leq 0$ **do**

$\mathbf{v} \leftarrow \mathbf{v} + \mathbf{x}_i^*;$

end

end

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

- ▶ Obsahem zprávy bude
 - ▶ stručný popis algoritmu,
 - ▶ klasifikační úspěšnost lineárního klasifikátoru a lineárního klasifikátoru s rozšířenou bází pro oba datasety,
 - ▶ Vizualizace trénovací a testovací chyby pomocí rozhodovací hranice klasifikátoru – viz přiložená funkce `linearplot11`.
 - ▶ váš komentář ke všem předchozím bodům.