Vytěžování dat, cvičení 11: Lineární klasifikátor

Miroslav Čepek, Michael Anděl

EVROPSKÁ UNIE

Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta elektrotechnická, ČVUT

(Binary) Bayesian Classification – Revision

- **Given** the data $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \ldots, (\mathbf{x}_n, y_n)\}$ with $y_i \in \{0, 1\}$, $\mathbf{x}_i \in \mathbb{X}$ where $\mathbb{X}$ is a feature space
- **Find** an *optimal* classification rule $r^* : \mathbb{X} \to \{0, 1\}$

What about aposteriori probability?

- $r^* = \mathbf{x} \mapsto 1 \text{ if } p\left(y = 1|\mathbf{x}\right) > p\left(y = 0|\mathbf{x}\right) \text{ else } 0$
- E.g.: The probability that a person y is a *woman*, if we see its height is $x = 160$ cm is higher, than the probability that it is a *man*.

But how to represent $p(y|x)$?

- $p(y|x) = \frac{p(x,y)}{p(x)}$
- $p(y = 1|\mathbf{x}) > p(y = 0|\mathbf{x}) \implies p(\mathbf{x}, y = 1) > p(\mathbf{x}, y = 0)$

But what about $p(x, y)$?

- $p(x, y) = p(x|y)p(y)$
- $p(1|\mathbf{x}) > p(0|\mathbf{x}) \implies p(x|1)p(1) > p(x|0)p(0)$
- $p(x|y = c) \approx p(x|\mu_c, \sigma_c^2)$ under the assumption that $x \sim \mathcal{N}(\mu_c, \sigma_c^2)$

$$\implies r^* \equiv \mathbf{x} \mapsto 1 \text{ if } p(1)p(x|\mu_1, \sigma_1^2) > p(0)p(x|\mu_0, \sigma_0^2) \text{ else } 0$$

(Binary) Bayesian Classification – Revision

Classification according to maximum likelihood:

$$r^* \equiv \mathbf{x} \mapsto 1 \text{ if } p(1)p(x|\mu_1, \sigma_1^2) > p(0)p(x|\mu_0, \sigma_0^2) \text{ else } 0$$

- $p(x|\mu_c, \sigma_C^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
- $J_c = \{i : 0 < i < n \wedge y_i = c\}$
- $p(c) \approx |J_c|$
- $\mu_c \approx \frac{1}{|J_c|} \sum_{j \in J_c} x_j$
- $\sigma_c^2 \approx \sum_{j \in J_c} (x_j - \mu_c)^2$

Generally:

$$r^* \equiv \mathbf{x} \mapsto \arg \max_{c \in \{0,1\}} p(c)p(\mathbf{x}|c)$$

- However, we need not to learn $p(0), p(1), \mu_0, \sigma_0^2, \ldots$
- instead, we can learn an easier function r directly from the experience, such that $r \approx r^*$. E.g.:

$$r = \mathbf{x} \mapsto sign(\mathbf{xw} + b)$$

How, do we get $\mathbf{w}$, b ?

Solving a Linear Separable Problem (Perceptron)

Let $\mathbf{x}_i \in \mathbb{R}^m$ be the i-th example from the m-dimensional space, and y_i be its class indicator. Finding a linear classifier (with *zero-training error*) means to solve the following system of equations for $\mathbf{w}$ and b:

$$\mathbf{x}_i\mathbf{w} + b \geq 0, \forall i : y_i = 1,$$
$$\mathbf{x}_i\mathbf{w} + b < 0, \forall i : y_i = 2.$$

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. $\mathbf{x}'_i\mathbf{v} > 0, \forall i$; using the *perceptron* algorithm as follows:

Data: $\{\mathbf{x}'_i\}_{i=1}^N$ transformed data samples
Result: $\mathbf{v} \in \mathbb{R}^m$ weights of the classifier
$\mathbf{v} \leftarrow 0$;
while $\exists \mathbf{x}_i : \mathbf{x}'_i\mathbf{v} \leq 0$ **do**
  $\mathbf{v} \leftarrow \mathbf{v} + \mathbf{x}'_i$;
end

Zadání domácího úkolu

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

Report

Váš report by měl obshovat:

- klasifikační úspěšnost lineárního klasifikátoru a lineárního klasifikátoru s rozšířenou bází pro oba datasety, tj. celkem 4 čísla
- vizualizace testovací chyby pomocí rozhodovací hranice klasifikátoru (viz přiložená funkce linearplot11) pro oba datasety. Toto pro lineární klasifikátor a lineární klasifikátor s rozšířenou bází, pro oba datasety. Tj. celkem 4 grafy
- váš komentář ke všem předchozím bodům.
- přiložený m-file