Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Thomas Seidl
Julian Busch, Evgeniy Faerman,
Florian Richter, Klaus Schmid

Knowledge Discovery in Databases

SS 2016

Exercise 10: SVM and Decision Trees

Regarding tutorials on 29.06.-01.07.2016.

Exercise 10-1 Support Vector Machines

Consider the following training data:

$$x_1 = (2, 3), x_2 = (3, 2), x_3 = (4, 4), x_4 = (4, 2)$$

$$x_5 = (6, 4), x_6 = (6, 3), x_7 = (7, 2), x_8 = (8, 3)$$

Let $y_A = \{-1\}, y_B = \{+1\}$ be the class indicators for both classes

$$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}.$$

(a) Plot the points and specify which of the points should be identified as support vectors.

(b) Draw the maximum margin line which separates the classes (you don't have to do any computations here). Write down the normalized normal vector $w' \in \mathbb{R}^2$ of the separating line and the offset parameter $b \in \mathbb{R}$.

(c) Consider the decision rule: $h(x) = \langle w', x \rangle + b$. Explain how this equation classifies points on either side of a line. Determine the class for the points $x_9 = (3, 4)$, $x_{10} = (7, 4)$ and $x_{11} = (5, 5)$.

Exercise 10-2 Kernel Trick

Consider the polynomial kernel function

$$K : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto (x^T y + \gamma)^p, \text{ with } p = 2, \gamma = 1.$$

Furthermore let

$$\phi : \mathbb{R}^2 \times \mathbb{R}^6, x \mapsto (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2).$$

Show that $K(x, y) = \langle \phi(x), \phi(y) \rangle$.

Exercise 10-3 Decision Trees

If you have problems finding appropriate gifts for parties, use a decision tree based on a dataset like the following:

gift ID	gender of the recipient	useful	beautiful	self-made	eatable	liked by the recipient
1	male	yes	yes	no	yes	yes
2	male	yes	yes	yes	no	yes
3	male	yes	no	yes	no	yes
4	male	yes	yes	no	no	yes
5	male	no	yes	yes	yes	yes
6	male	no	little	no	yes	yes
7	male	no	no	yes	no	no
8	male	no	yes	no	no	no
9	female	no	yes	no	yes	yes
10	female	no	yes	no	no	yes
11	female	no	little	yes	no	yes
12	female	no	no	no	no	yes
13	female	no	little	yes	no	yes
14	female	no	little	yes	no	yes
15	female	no	little	no	yes	no
16	female	no	little	no	no	no

(a) Generate the decision tree by using the Gini-index.

(b) Generate the decision tree by using the Information Gain.