Johannes Gutenberg-Universität Mainz
Institut für Informatik
Data Mining
Prof. Dr. Stefan Kramer

Machine Learning - Sheet 9

26.06.2017
Deadline: 03.07.2017 - 14:00

Task 1: SVM with ∞-Dimensional Feature Space *(5 Points)*

Consider the attached script `ml17-exercise-09-spiral.py`. For each *slope* $\in \{0.1, 1, 5\}$ adjust the parameter σ so that the two classes are properly separated. Document your results in a PDF file (with plots and parameter settings).

Task 2: Maximum Margin Hyperplane *(8 Points)*

The goal of this exercise is to get acquainted with the notation used in the lecture. We shall derive an explicit formula for the maximum margin hyperplane for a particularly simple case where we have just two support vectors.

Recall that the feature space of an SVM kernel can be an arbitrarily complex (not necessarily finite-dimensional) vector space equipped with an operation that behaves just like the dot product in $\mathbb{R}^n$. Let V be some $\mathbb{R}$-vector space (that means, we can add and subtract elements of V, and scale them with real factors). An *inner product* on V is a two-argument function $\langle \bullet, \bullet \rangle : V \times V \to \mathbb{R}$ that satisfies the following conditions:

- Symmetry: for all $x, y \in V$ it holds: $\langle x, y \rangle = \langle y, x \rangle$.
- Bilinearity: for all $x, y, z \in V$ and $\alpha, \beta \in \mathbb{R}$ it holds:

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \qquad \langle z, \alpha x + \beta y \rangle = \alpha \langle z, x \rangle + \beta \langle z, y \rangle$$

- Positive semidefiniteness: for all $x \in V$ it holds $\langle x, x \rangle \geq 0$ with equality only for $x = 0$.

We call $x, y \in V$ *orthogonal* iff $\langle x, y \rangle = 0$ holds. We define the *norm* of an element $v \in V$ as follows:

$$\|v\| := \sqrt{\langle v, v \rangle}.$$

Notice that $\|v\|^2 = \langle v, v \rangle$ always holds.

(1) Check that the following equations hold for all $a, b \in V$:

$$\|a + b\|^2 = \|a\|^2 + 2 \langle a, b \rangle + \|b\|^2 \qquad \|a - b\|^2 = \|a\|^2 - 2 \langle a, b \rangle + \|b\|^2 .$$

Hint: rewrite squared norm as inner product, apply bilinearity as long as possible, use symmetry.

(2) Let $a, b \in V$ be two orthogonal vectors. Set $c := a - b$. Show the *Pythagorean theorem*:

$$\|c\|^2 = \|a\|^2 + \|b\|^2 .$$

Hint: Use (1), use definition of orthogonality.

Now, let $x, y \in V$ be two support vectors, with class of x being $+1$ and class of y being -1. Assume that $x \neq y$. Suppose that we want to find $w \in V$ and $b \in \mathbb{R}$ such that

$$\begin{aligned} \langle w, x\rangle + b &= +1 \qquad (*) \\ \langle w, y\rangle + b &= -1 \end{aligned}$$

holds, and moreover, we want to minimize $\|w\|$.

(3) Show that for any w, b (not necessarily optimal) that satisfy equations $(*)$ it must hold:

$$b = -\frac{\langle w, x\rangle + \langle w, y\rangle}{2},$$

that is, every solution w determines a corresponding $b \equiv b(w)$.

(4) Check that $\hat{w}, b$ with $\hat{w} := \frac{2}{\|x-y\|^2}(x - y)$ and $b \equiv b(\hat{w})$ satisfy both equations in $(*)$.
Hint: Expand definition of $\hat{w}$ as late as possible.

(5) Show that any other $w \in V, b \in \mathbb{R}$ that satisfy equations $(*)$ must also satisfy $\langle w, x - y\rangle = 2$.
Hint: Use bilinearity, "add 0", use the two equations $(*)$.

(6) Let $w \in V$ satisfy $\langle w, x - y\rangle = 2$. Show: $\|w\|^2 = \|\hat{w}\|^2 + \|w - \hat{w}\|^2$.
Hint: Check that $\hat{w}$ and $(w - \hat{w})$ are orthogonal, "add zero", summon Pythagoras!

(7) Conclude: For any other $w \neq \hat{w}$ that satisfies $(*)$ it holds: $\|w\| > \|\hat{w}\|$, that is: $\hat{w}$ is the optimal solution.

(8) Now let $V = \mathbb{R}^2$, $\langle x, y\rangle = x_1y_1 + x_2y_2$. Calculate $\hat{w}$, $b \equiv b(\hat{w})$ for support vectors $x = (5, 5)$ and $y = (2, 1)$, sketch the decision boundary.

Task 3: String Spectrum Kernel *(7 Points)*

Let Σ be a finite alphabet. Let Σ^p be the set of strings of length p and $\Sigma^* := \bigcup_{p=0}^{\infty} \Sigma^p$ the set of all strings with characters from Σ. For $p \in \mathbb{N}$ we define the *p-spectrum ϕ^p of a string $s \in \Sigma^*$* as follows:

$$\phi^p(s) := (occ(s, u))_{u \in \Sigma^p} \qquad occ(s, u) := \#\{(\pi, \sigma) \in \Sigma^* \times \Sigma^* \mid \pi u \sigma = s\},$$

that is, $occ(s, u)$ counts how many times u occurs in s, and the p-spectrum of string s contains the counts for all substrings u of length p. We define the p-spectrum kernel $\kappa_p : \Sigma^* \times \Sigma^* \to \mathbb{R}$ as follows:

$$\kappa_p(s, t) := \sum_{u \in \Sigma^p} occ(s, u) \cdot occ(t, u).$$

(1) What is the input space? What is the feature space? What is the mapping from input space to feature space? Write κ_p in such a way that it becomes obvious that it is indeed a kernel.

(2) What is the dimension of the feature space?

(3) Consider the special case where $\Sigma = \{A, T, G, C\}$ and $p = 2$. Let $s = ATATGCTGCA$, $t = CGATGCTGCATG$. Compute 2-spectra for s and t, compute $\kappa_2(s, t)$.
Hint: Choose an appropriate way to write down $\phi^2(s)$ and $\phi^2(t)$. A DIN-A4 page is roughly 21cm wide, this is enough for a 16cm wide table with two rows.