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

Machine Learning - Sheet 8

21.06.2018
Deadline: 29.06.2018 - 16:00

Task 1: Maximum Margin Hyperplane - Dimensionality *(3 Points)*
Show that, irrespective of the dimensionality of the data space, a data set consisting of just two data points, one from each class, is sufficient to determine the location of the maximum-margin hyperplane.

Task 2: Maximum Margin Hyperplane - Constraint *(4 Points)*
Show that, if the 1 on the right-hand side of the constraint $y_n(\mathbf{w}^T\Phi(\mathbf{x}_n) + b) \geq 1$, $n = 1, \ldots, N$, is replaced by some arbitrary constant $\gamma > 0$, the solution for the maximum margin hyperplane is unchanged.

Task 3: Maximum Margin Hyperplane - Margin *(4 Points)*
Show that the value ρ of the margin for the maximum-margin hyperplane is given by

$$\frac{1}{\rho^2} = \sum_{n=1}^{N} a_n,$$

where $\{a_n\}$ are given by maximizing

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m k(\mathbf{x}_n, \mathbf{x}_m)$$

subject to the constraints $a_n \geq 0$, $n = 1, \ldots, N$, and $\sum_{n=1}^{N} a_n y_n = 0$.

Task 4: Kernels *(9 Points)*
For any non-empty set $\mathcal{X}$ a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be *positive semi-definite*, if the following conditions hold for all $x_1, \ldots, x_m \in \mathcal{X}$:

- $k(x_i, x_j) = k(x_j, x_i)$ for all x_i, x_j (symmetry)
- $\forall c_1, \ldots, c_m \in \mathbb{R}$: $\sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) \geq 0$

Every positive semi-definite kernel can be represented as a dot product in a linear space, thus allowing for the *kernel trick*.

(a) Show that the dot product $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, k(x, y) := \sum_{i=1}^{n} x_i y_i$ is a positive semi-definite kernel.

(b) Show that the polynomial kernel $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, k(x, y) := (\sum_{i=1}^{n} x_i y_i)^2$ is a positive semi-definite kernel.

Now, we want to build more complex kernels from simpler ones. Suppose that *we already know* that if $k, k_1, k_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are kernels, then

- $k_1 + k_2$ (i.e, $(x, y) \mapsto k_1(x, y) + k_2(x, y)$)
- $k_1 \cdot k_2$ (i.e, $(x, y) \mapsto k_1(x, y) \cdot k_2(x, y)$)
- $\exp \circ k$ (i.e, $(x, y) \mapsto \exp(k(x, y))$)

are also kernels. Relying only on the definition of the kernel and these three "rules", show that the following functions are also kernels:

(c) Let $d \in \mathbb{N}$ be some exponent, and k a kernel. Show that k^d, i.e. $(x, y) \mapsto (k(x, y))^d$ is a kernel.

(d) Let $n \in \mathbb{N}$, $c_0, \ldots, c_n \in \mathbb{R}_{\geq 0}$, let k be a kernel. Show that $\sum_{i=0}^{n} c_i \cdot k^i$ is also a kernel.

(e) Let $f : \mathcal{X} \to \mathbb{R}$ an arbitrary function. Show that $(x, y) \mapsto f(x)k(x, y)f(y)$ is a kernel.