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

Machine Learning - Sheet 10

03.07.2017
Deadline: 10.07.2017 - 23:55

Task 1: Kernels and Positive Semidefinite Matrices *(3 Points)*

The goal of this exercise is to understand how the characterization of kernels through positive semidefinite matrices helps us recognize certain compositions of kernels as kernels.

Recall that a function $k : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ is a kernel if and only if for all $n \in \mathbb{N}$ and all choices $x_1, \ldots, x_n \in \mathfrak{X}$ the matrix $K := (k(x_i, x_j))_{i,j=1}^n$ is positive semidefinite, that is, for all $\alpha \in \mathbb{R}^n$ it holds

$$\alpha^T K \alpha \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i k(x_i, x_j) \alpha_j \geq 0.$$

Using this characterization, show that if $k_1, k_2 : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ are two kernels, then their pointwise sum $k_1 + k_2$ (i.e., $(x, y) \mapsto k_1(x, y) + k_2(x, y)$) is again a kernel.

Task 2: Compositions of Kernels *(10 Points)*

We want to build more complex kernels from simpler ones. Suppose that *we already know* that if $k, k_1, k_2 : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$ 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, and encode your proofs as Java programs in the provided file `Kernels.java`.

(1) Let $c \in \mathbb{R}$, $c \geq 0$. Then the constant function $\mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$, $(x, y) \mapsto c$ is a kernel.

(2) Let $c \geq 0$ be as above, and k be a kernel. Show that $c \cdot k$ is also a kernel.

(3) 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.

(4) 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.

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

(6) Show that the Gaussian radial basis function $(x, y) \mapsto \exp\left(- \|x - y\|^2 / (2\sigma^2)\right)$ is a kernel.
Hint: Use $\|x - y\|^2 = \|x\|^2 - 2 \cdot \langle x, y \rangle + \|y\|^2$, then use (5) and the rule with the exponential function. You will probably also need (2) and the actual definition of the kernel.

(7) Show that for $\mathfrak{X} = \mathbb{R}^n$ and all $d \in \mathbb{N}$ the function $(x, y) \mapsto (\langle x, y \rangle + 1)^d$ is a kernel.

Test your implementation using the Java compiler (the type-checker ensures that the compositions are indeed kernels), and the provided tests (they ensure that your kernels actually compute the right values).

Task 3: Converging Connections and D-Separation *(7 Points)*

Consider the Bayesian network for three Boolean random variables A, B and V in Figure 1. In the following, we abbreviate the events $\{X = true\}$ and $\{X = false\}$ for a Boolean random variable X by just "X" and "$\neg X$" respectively.

Suppose that A and B are independent with $\mathbb{P}[A] = 0.25$, $\mathbb{P}[B] = 0.5$, and that the conditional probabilities $\mathbb{P}[V = v \,|\, A = a, B = b]$ for Boolean a, b, v are given by the following table:

	A, B	$A, \neg B$	$\neg A, B$	$\neg A, \neg B$
V	0.1	0.9	0.9	0.1
$\neg V$	0.9	0.1	0.1	0.9

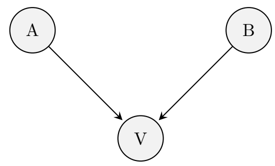

Figure 1: Bayesian net.

(1) Explicitly write out the joint probability distribution of A, B, V defined by this network, that is, for all combinations of $a, b, v \in \{true, false\}$ compute $\mathbb{P}[A = a, B = b, V = v]$.

(2) Using the table from (1), compute $\mathbb{P}[A, B]$, $\mathbb{P}[A]$, $\mathbb{P}[B]$, $\mathbb{P}[A, B \,|\, V]$, $\mathbb{P}[A \,|\, V]$, $\mathbb{P}[B \,|\, V]$.

(3) Compare $\mathbb{P}[A, B]$ with $\mathbb{P}[A] \cdot \mathbb{P}[B]$, and $\mathbb{P}[A, B \,|\, V]$ with $\mathbb{P}[A \,|\, V] \cdot \mathbb{P}[B \,|\, V]$.

(4) Are A, B independent? Are A, B conditionally independent given V?

(5) When are A and B d-separated, when are they not d-separated?