TU München
Fakultät für Informatik
PD Dr. Rudolph Triebel
John Chiotellis

Machine Learning for Robotics and Computer Vision
Winter term 2015

Solution Sheet 4

Topic: Boosting and Kernels
December 18th, 2015

Exercise 1: Adaboost

See code

Exercise 2: Kernels

Remember that for a function to be a valid kernel, it must correspond to a scalar product in some (perhaps infinite dimensional) feature space. First let us write down the kernel constructing rules that we will use: Let K_1 and K_2 be kernels on $\mathcal{X} \subseteq \mathbb{R}^n$ and K_3 be kernel on $f : \mathcal{X} \to \mathbb{R}^m$.

Rules

1. $K(x, y) = K_1(x, y) + K_2(x, y)$
2. $K(x, y) = cK_1(x, y) \quad , c > 0$
3. $K(x, y) = K_1(x, y)K_2(x, y)$
4. $K(x, y) = K_3(f(x), f(y))$
5. $K(x, y) = x^T By \quad$, for B square, symmetric and positive semi-definite
6. $K(x, y) = c \quad , c > 0$

Proofs

1.

$$\begin{aligned} K_1(x, y) + K_2(x, y) &= \phi_1(x)^T \phi_1(y) + \phi_2(x)^T \phi_2(y) \\ &= (\phi_1(x) \quad \phi_2(x))^T (\phi_1(y) \quad \phi_2(y)) \\ &= \phi(x)^T \phi(y) \end{aligned}$$

with $\phi(x) = (\phi_1(x) \quad \phi_2(x))^T$.

2.

$$\begin{aligned} cK_1(x,y) &= c\phi_1(x)^T\phi_1(y) \\ &= \sqrt{c}\sqrt{c}\phi_1(x)^T\phi_1(y) \\ &= (\sqrt{c}\phi_1(x))^T(\sqrt{c}\phi_1(y)) \\ &= \phi(x)^T\phi(y) \end{aligned}$$

with $\phi(x) = (\frac{\sqrt{c}}{\sqrt{n_1}}\phi_1(x)_1, \dots, \frac{\sqrt{c}}{\sqrt{n_1}}\phi_1(x)_{n_1})^T$

and $\phi_1(x) \in \mathbb{R}^{n_1}$, $\phi_2(x) \in \mathbb{R}^{n_2}$, $\phi(x) \in \mathbb{R}^{n_1+n_2}$.

3.

$$\begin{aligned} K_1(x,y)K_2(x,y) &= \phi_1(x)^T\phi_1(y)\phi_2(x)^T\phi_2(y) \\ &= (\sum_i \phi_1(x)_i\phi_1(y)_i)(\sum_j \phi_2(x)_j\phi_2(y)_j) \\ &= \sum_i\sum_j \phi_1(x)_i\phi_1(y)_i\phi_2(x)_j\phi_2(y)_j \\ &= \sum_i\sum_j \phi_1(x)_i\phi_2(x)_j\phi_1(y)_i\phi_2(y)_j \\ &= \sum_k \phi_k(x)\phi_k(y) \\ &= \phi(x)^T\phi(y) \end{aligned}$$

with $\phi(x) = \begin{pmatrix} \phi_1(x)_1\phi_2(x)_1 \\ \vdots \\ \phi_1(x)_1\phi_2(x)_{n_2} \\ \phi_1(x)_2\phi_2(x)_1 \\ \vdots \\ \phi_1(x)_{n_1}\phi_2(x)_{n_2} \end{pmatrix} \in \mathbb{R}^{n_1 \cdot n_2}$.

4. Since K_3 is a valid kernel in $\mathbb{R}^m$ there is a feature space ψ for which it holds

$$K_3(f(x), f(y)) = \psi(f(x))^T\psi(f(y))$$

Therefore it is also a valid kernel in $\mathbb{R}^n$ with with $\phi(x) = \psi(f(x))$.

5. Since B is symmetric and positive definite we can use its Cholesky decomposition:

$$x^TBy = x^TLL^Ty = (L^Tx)^T(L^Ty) = \phi(x)^T\phi(y)$$

with $\phi(x) = L^Tx$.

6. c is a kernel (Rule 5 with $B = I$ and Rule 4 with $\phi(x) = \phi(y) = (\frac{\sqrt{c}}{\sqrt{m}}, \dots, \frac{\sqrt{c}}{\sqrt{m}})^T$ if $\phi(x) \in \mathbb{R}^m$.

Gaussian Kernel First let us prove that the exponential of a kernel is also a kernel. Using the Taylor expansion of the exponential, we have:

$$exp(K_1(x,y)) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} K_1(x,y)^n$$

Using the rules we defined, we see that

- $K_1(x,y)^n$ is a kernel (iteratively Rule 3)
- $(\frac{1}{n!})K_1(x,y)^n$ is a kernel (Rule 2)
- $\sum_{n=1}^{\infty}(\frac{1}{n!})K_1(x,y)^n$ is a kernel (iteratively Rule 1)
- 1 is a kernel (Rule 6)
- the whole expression is a kernel because of Rule 1.

Now we can rewrite the Gaussian kernel as follows:

$$\begin{aligned}\exp(-\frac{|x-y|^2}{2\sigma^2}) = \exp(-\frac{(x-y)^T(x-y)}{2\sigma^2}) &= \exp(-\frac{x^Tx - 2x^Ty + y^Ty}{2\sigma^2}) \\ &= \exp(-\frac{x^Tx}{2\sigma^2})\exp(-\frac{y^Ty}{2\sigma^2})\exp(\frac{x^Ty}{2\sigma^2})\end{aligned}$$

The expression $\frac{x^Ty}{2\sigma^2}$ is a kernel because of Rule 5 ($B = I$) and Rule 2 ($c = \frac{1}{2\sigma^2}$) and as we showed its exponential is also a kernel.

The remaining expression is a kernel because of Rule 4 with $\phi(x) = (\frac{\exp(-\frac{x^Tx}{2\sigma^2})}{\sqrt{m}}, \ldots, \frac{\exp(-\frac{x^Tx}{2\sigma^2})}{\sqrt{m}})$ and $\phi(y) = (\frac{\exp(-\frac{y^Ty}{2\sigma^2})}{\sqrt{m}}, \ldots, \frac{\exp(-\frac{y^Ty}{2\sigma^2})}{\sqrt{m}})$.

Polynomial Kernel The polynomial kernel is defined as

$$K(x,y) = (x^Ty + c)^d \quad , c > 0, d \in \mathbb{N}$$

Using the rules we defined we see that:

- x^Ty is a kernel (Rule 5 with $B = I$)
- c is a kernel (Rule 6)
- $x^Ty + c$ is a kernel (Rule 1)
- $(x^Ty + c)^d$ is a kernel (Rule 3)