TU München
Fakultät für Informatik
Dr. Rudolph Triebel
Matthias Vestner

Machine Learning for Robotics and Computer Vision
Winter term 2013

Solution Sheet 4
Topic: Boosting
December 6th

Exercise 1: Constructing kernels

During this solution we assume the feature spaces of k_1 and k_2 to have finite dimensions. Thus they can be written as $k_1(x_1, x_2) = \phi_1(x_1)^T \phi_1(x_2)$, $k_2(x_1, x_2) = \phi_2(x_1)^T \phi_2(x_2)$, where $\phi_1(x) \in \mathbb{R}^{n_1}$, $\phi_2(x) \in \mathbb{R}^{n_2}$. Note however that in general feature spaces can be infinite dimensional (e.g. $\phi(x) \in l^2(\mathbb{R})$, see 4.). We now have to define new kernels via a scalarproduct $k(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$

1. To warm up:
   $\phi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix} \in \mathbb{R}^{n_1+n_2}$

2. Note that the matrix-products do not commute, so it is a bit of work:

$$\begin{aligned}
k(x_1, x_2) &= \phi_1(x_1)^T \phi_1(x_2) \phi_2(x_1)^T \phi_2(x_2) \\
&= (\sum_i (\phi_1(x_1))_i (\phi_1(x_2))_i)(\sum_j (\phi_2(x_1))_j (\phi_2(x_2))_j) \\
&= \sum_i \sum_j (\phi_1(x_1))_i (\phi_1(x_2))_i (\phi_2(x_1))_j (\phi_2(x_2))_j \\
&= \underbrace{\sum_i \sum_j}_{\sum_k} \underbrace{(\phi_1(x_1))_i (\phi_2(x_1))_j}_{\phi_k(x_1)} \underbrace{(\phi_1(x_2))_i (\phi_2(x_2))_j}_{\phi_k(x_2)} \\
\Rightarrow \phi(x) &= \begin{pmatrix} (\phi_1(x_1))_1 (\phi_2(x_1))_1 \\ \vdots \\ (\phi_1(x_1))_1 (\phi_2(x_1))_{n_2} \\ (\phi_1(x_1))_2 (\phi_2(x_1))_1 \\ \vdots \\ (\phi_1(x_1))_{n_1} (\phi_2(x_1))_{n_2} \end{pmatrix} \in \mathbb{R}^{n_1 \cdot n_2}
\end{aligned}$$

3. $\phi(x) = f(x)\phi_1(x)$

4. Again we write the scalarproduct as a sum:

$$\begin{aligned}\exp((\phi_1(x))^T\phi(y)) &= \exp(\sum(\phi_1(x))_i(\phi_1(y))_i) \\ &= \prod \exp((\phi_1(x))_i(\phi_1(y))_i)\end{aligned}$$

   Since we already know that the product of kernels is again a kernel it remains to show, that $\exp((\phi(x))_i(\phi(y))_i)$ is a kernel for a fixed index i. In the following we will omit i and imagine ϕ_1 to be a scalar-valued function

$$\exp(\phi_1(x))(\phi_1(y)) = \sum_{k=0}^{\infty}\frac{1}{k!}(\phi_1(x))^k(\phi_1(y))^k$$

   This is an inner product in $l^2(\mathbb{R})$ with

$$\phi(x) = \begin{pmatrix}\phi_1(x) \\ \frac{1}{\sqrt{2}}\phi_1(x)^2 \\ \frac{1}{\sqrt{6}}\phi_1(x)^3 \\ \vdots \\ \frac{1}{\sqrt{k!}}\phi_1(x)^k \\ \vdots\end{pmatrix}$$

5. $\phi(x) = \sqrt{A}x$, $(A = Q^T\Lambda Q \Rightarrow \sqrt{A} = Q^T\sqrt{\Lambda}Q,\ (\sqrt{A})^T\sqrt{A} = A)$

Exercise 2: Polynomial kernel

1. $d = 1$: $\phi(x) = x$. Inductionstep: Exercise 1, 2.

2. Consider first $d = 2$:

$$\begin{aligned}(x_i^Tx_j)^2 &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= x_{i1}^2x_{j1}^2 + 2x_{i1}x_{j1}x_{i2}x_{j2} + x_{i2}^2x_{j2}^2 \\ \phi(x) &= \begin{pmatrix}x_1^2 & \sqrt{2}x_1x_2 & x_2^2\end{pmatrix}^T\end{aligned}$$

   For larger d the coefficients can be obtained by using the Binomial theorem/Pascal's triangle:

$$\begin{array}{ccccccccc} &&&& 1 &&&& \\ &&& 1 && 1 &&& \\ && 1 && 2 && 1 && \\ & 1 && 3 && 3 && 1 & \\ 1 && 4 && 6 && 4 && 1\end{array}$$

3.

$$\begin{aligned}\tilde{k}_2(x_i, x_j) &= (x_i^Tx_j + d)^2 = (x_i^Tx_j)^2 + 2dx_i^Tx_j + d^2 \\ \phi(x) &= \begin{pmatrix}x_1^2 & \sqrt{2}x_1x_2 & x_2^2 & \sqrt{2d}x_1 & \sqrt{2d}x_2 & d\end{pmatrix}^T\end{aligned}$$

Exercise 3: Programming

```
function [above] = feature_test(X,n,c,phi)
% FEATURE_TEST returns a boolean vector, indicating on whether the features
% of X lies above a hyperplane
%   X: matrix of elements in the original space, e.g. R^2
%   n: normal of the hyperplane in the feature-space, e.g. R^3
%   c: distance of the hyperplane to the origin, scalar
%   phi: function-handle for the mapping from the original space to the
%   feature-space, e.g. R^2 --> R^3

psi = phi(X(:,1),X(:,2));
above = (psi*n>c);
end
```