ASSIGNMENT 6: KERNELS

Institute for Machine Learning

JOHANNES KEPLER
UNIVERSITY LINZ

Institute for
Machine Learning

Contact

Heads: Johannes Kofler,
Markus Holzleitner

Institute for Machine Learning
Johannes Kepler University
Altenberger Str. 69
A-4040 Linz

E-Mail: `theoretical@ml.jku.at`

Only mails to this list are answered!

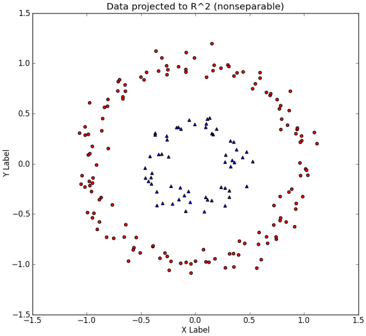
Institute Homepage

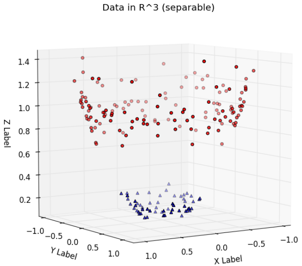
Copyright statement:

Agenda:

- Kernels: Intuition+main idea
- Kernels: Mathematical formalism
- Examples

Kernel trick: Intuition





Kernels: Basic idea (1)

- Basic idea: transform data into a higher-dimensional space such that problem hopefully becomes linearly separable there.
- More formal: choose a Hilbert space $\mathcal{H}$ and a (nonlinear) mapping $\Phi : X \to \mathcal{H}$.
- Then try to apply linear method presented previously in the space $\mathcal{H}$.
- Problem: how to specify $\mathcal{H}$ and Φ?

Kernels: Basic idea (2)

- In many machine learning tasks: only scalar products of pairs of samples matter. Therefore: not necessary to explicitly know $\mathcal{H}$ and Φ.
- Only need $\Phi(\mathbf{x}^i) \cdot \Phi(\mathbf{x}^j)$ for all $\mathbf{x}^i, \mathbf{x}^j$ $(i, j = 1, \ldots, l)$.
- Required for computing the classification of a new sample $\mathbf{x}$: $\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}^i)$ for all $i = 1, \ldots, l$.

Kernels: more formal

- Suppose we are given a mapping $k : X \times X \to \mathbb{R}$ (the kernel) for which we know that there exists Hilbert space $\mathcal{H}$ and mapping $\Phi : X \to \mathcal{H}$ with $k(\mathbf{x}^i, \mathbf{x}^j) = \Phi(\mathbf{x}^i) \cdot \Phi(\mathbf{x}^j)$ for all $\mathbf{x}^i \in X$, $i = 1, ..., l$
- This is the case $\Leftrightarrow$ k is positive semi-definite and symmetric, i.e.
  1. $\sum_{i,j} c_i k(\mathbf{x}^i, \mathbf{x}^j) c_j \geq 0$
  2. $k(\mathbf{x}^i, \mathbf{x}^j) = k(\mathbf{x}^j, \mathbf{x}^i)$

  for $i = 1, ..., l$, $c_i \in \mathbb{R}$, $\mathbf{x}^i \in X$.
- Equivalent formulation: Gram matrix $\mathbf{K} = (k_{ij})_{i=1,...,l}^{j=1...,l} = (k(\mathbf{x}^i, \mathbf{x}^j))_{i=1,...,l}^{j=1...,l}$ is positive semi-definite and symmetric.
- In practice: make an a priori choice of k using common sense and, if available, prior knowledge about problem: $\to$ "kernel trick".

How to obtain large class of kernels

- Which kernels? $\rightarrow$ Mercer's theorem: The following statements are equivalent:
  - $k : X^2 \rightarrow \mathbb{R}$ can be written as $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$ for some Hilbert space $\mathcal{H}$ and $\Phi : X \rightarrow \mathcal{H}$.
  - The inequality
    $$\int_{X^2} k(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) f(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0$$
    holds for all square-integrable functions $f \in L^2(X)$.
- More details with proofs: e.g. these notes, chapter 10.5.
- Standard kernels:
  1. Linear: $k(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
  2. Polynomial: $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + \sigma^2)^d$
  3. Gaussian/RBF: $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{y}\|^2\right)$
  4. Sigmoid: $k(\mathbf{x}, \mathbf{y}) = \tanh(\alpha \mathbf{x} \cdot \mathbf{y} + \beta)$

Further information

- The sigmoid kernel is not a very popular choice; moreover, it is not positive semi-definite for all choices of α and β.
- RBF-kernel:
  1. Most popular choice
  2. Maps into a hyper-sphere of radius 1.
  3. Hilbert space corresponding to RBF kernel is infinitely dimensional.
- How to construct kernels in real-world applications?
  1. If we can define $\mathcal{H}$ (most often $\mathbb{R}^k$) and Φ explicitly $\rightarrow$ done!
  2. Products, weighted sums,... applied to positive semi-definite kernels give semi-definite kernels.
  3. Suppose that we have a mapping $\Psi : X \rightarrow Y$, where Y is some feature space, and a semi-definite kernel $k : Y^2 \rightarrow \mathbb{R}$. Then $k' : X^2 \rightarrow \mathbb{R}$, defined as $k'(\mathbf{x}, \mathbf{y}) = k(\Psi(\mathbf{x}), \Psi(\mathbf{y}))$ is also a positive semi-definite kernel.