IML Tutorial 4: Kernels

Mikhail Karasikov

11 March 2019

Motivation

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_H,$$

Why use kernels:

- For efficient feature transformation (basis expansion)
  Example: use polynomial kernel instead of computing all monomials from features
- To use linear models with non-linearly separable data (through mapping the data to a higher dimensional space)
  Example: solving XOR with a single Perceptron
- Use with complex structured data to avoid computationally expensive feature extraction
  Example: String kernels, graph kernels

Kernels

Definition

A symmetric function $K: X \times X \to \mathbb{R}$ is called a kernel if

1. symmetric: $K(x, y) = K(y, x) \quad \forall x, y \in X$,
2. positive semi-definite: $[K(x_i, x_j)]_{i,j=1}^n \succeq 0 \quad \forall x_1, \ldots, x_n \in X$, $n < \infty$, that is,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(x_i, x_j) \geq 0 \quad \forall c_1, \ldots, c_n \in \mathbb{R}.$$

- Kernels compute inner products:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_H,$$

  where $\phi: X \to H$ is a *feature map* from X to a *Hilbert space* H (a vector space with an inner product inducing a complete metric space).
- Any inner product is positive definite, hence defines a Kernel

Kernels

Properties

- Linear combination: $\sum_{i=1}^{n} \alpha_i K_i(\cdot, \cdot)$ is a kernel if $\alpha_i \geq 0 \; \forall i$
- Product: $\prod_{i=1}^{n} K_i^{d_i}(\cdot, \cdot)$ is a kernel if $d_i \in \mathbb{N} \; \forall i$
- Limit: $\lim_{n \to \infty} K_n(\cdot, \cdot)$ is a kernel

Examples

- Linear: $\quad K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{y}$
- Polynomial: $\quad K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^T \boldsymbol{y} + r)^n, \quad r \geq 0,\ n \geq 1$
- Gaussian (RBF Kernel): $\quad K(\boldsymbol{x}, \boldsymbol{y}) = e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/h}, \quad h > 0$
- Laplacian kernel: $\quad K(\boldsymbol{x}, \boldsymbol{y}) = e^{-\alpha\|\boldsymbol{x}-\boldsymbol{y}\|_1}, \quad \alpha > 0$

How to construct/choose a kernel?

- Use domain knowledge
- Cross-validation

Exercises

Assume $K_1 : X_1^2 \to \mathbb{R}$ and $K_2 : X_2^2 \to \mathbb{R}$ are kernels

1. Show that the following functions are kernels
   1.1 $K_1(x_1, y_1) + K_2(x_2, y_2)$ is a kernel
   1.2 $K_1(x_1, y_1)K_2(x_2, y_2)$ is a kernel
   1.3 $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y}$
   1.4 $K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^T\boldsymbol{y} + r)^n, \quad r \geq 0, n \geq 1$
   1.5 $K(\boldsymbol{x}, \boldsymbol{y}) = e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2/h}, \quad h > 0$
2. Construct a feature map ϕ associated with the RBF Kernel
3. Construct a Kernelized Ridge Regression