Tutorial: Kernel Methods

Chih-Wei Chang B00201037
Chung-Yen Hung B00201015

November 27, 2014

Abstract

Here comes the Abstract

1 Introduction to Kerenl Method

1.1 What Is Kernel and Kernel Trick

1.2 Why Use Kernel and How Kernel Can Be Used

2 Common Kernel

2.1 Hilber Space and RKHS

Definition 2.1 (Inner Product Space)**.** An inner product space $\mathcal{X}$ is a vector space with an associated inner product $\langle \cdot, \cdot \rangle : \mathcal{X} \to \mathbb{R}$ that satisfies:

- Symmetry: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- Linearity: $\langle a \cdot \mathbf{x}, \mathbf{y} \rangle = a \cdot \langle \mathbf{x}, \mathbf{y} \rangle$ and $\langle \mathbf{w} + \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{w}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle$
- Positive Semi-Definiteness: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$

The inner product space is strict if $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ iff $\mathbf{x} = 0$

Definition 2.2 (Hilbert Space)**.** A strict inner product space $\mathcal{F}$ is a Hilbert space if it is:

- Complete: Every Cauchy sequence $\{h_i \in \mathcal{F}\}_{i=1}^{\infty}$ converges to an element $h \in \mathcal{F}$
- Separable: There is a countable subset $\hat{\mathcal{F}} = \{h_i \in \mathcal{F}\}_{i=1}^{\infty}$ such that for all $h \in \mathcal{F}$ and $\epsilon > 0$, there exists $h_i \in \hat{\mathcal{F}}$ such that $\|h_i - h\| < \epsilon$.

The interval $[0, 1]$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$ and Euclidean spaces $\mathbb{R}^D$ are all Hilber spaces.

2.2 Kerenl Function

2.3 Polynomial Kernel

The polynomial kernel is defined as

$$K_{\Phi_d}(x, y) = (\langle x, y\rangle + \alpha)^d$$

As a kernel, K corresponds to an inner product in a feature space based on some mapping:

$$K_{\Phi_d}(x, y) = \langle \Phi_d(x), \Phi_d(y)\rangle$$

Let $d = 2$, so we get the special case of the quadratic kernel

$$K(x, y) = (\sum_{i=1}^{n} x_i y_i + \alpha)^2 = \sum_{i=1}^{n}(x_i^2)(y_i^2) + \sum_{i=2}^{n}\sum_{j=1}^{i-1}(\sqrt{2}x_i x_j)(\sqrt{2}y_i y_j) + \sum_{i=1}^{n}(\sqrt{2c}x_i)(\sqrt{2c}y_i)$$

From this it follows that the feature map is given by:

$$\Phi_2(x) = \langle x_n^2, \ldots, x_1^2, \sqrt{2}x_n x_{n-1}, \ldots, \sqrt{2}x_n x_1, \sqrt{2}x_{n-1}x_{n-2}, \ldots, \sqrt{2c}x_n, \ldots, \sqrt{2c}x_1, c\rangle$$

2.4 Gaussian Kernel

3 Kernel Machines

3.1 Kernel PCA

3.2 Kernel SVM

3.3 Kerenl Ridge Regression

3.4 Kernel Logistic Regression

4 Output Kernel

4.1 Kernel in Output Space

4.2 PLST and CPLST

4.3 Other Possible Output Kernel Techniques

5 Conclusion

Here comes the Conclusion