10.10 Kernel functions

In the context of SVM's, a kernel function is a function of the form $K(\vec{x}, \vec{y})$, where $\vec{x}$ and $\vec{y}$ are n-dimensional vectors, having a special property. These functions are used to obtain SVM-like classifiers for two-class datasets which are not linearly separable.

10.10.1 Definition

Let $\vec{x}$ and $\vec{y}$ be arbitrary vectors in the n-dimensional vector space $\mathbb{R}^n$. Let ϕ be a mapping from $\mathbb{R}^n$ to some vector space. A function $K(\vec{x}, \vec{y})$ is called a kernel function if there is a function ϕ such that $K(\vec{x}, \vec{y}) = \phi(\vec{x}) \cdot \phi(\vec{y})$.

10.10.2 Examples

Example 1

Let

$$\begin{aligned}\vec{x} &= (x_1, x_2) \in \mathbb{R}^2\\ \vec{y} &= (y_1, y_2) \in \mathbb{R}^2\end{aligned}$$

We define

$$K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y})^2.$$

We show that this is a kernel function. To do this, we note that

$$\begin{aligned}K(\vec{x}, \vec{y}) &= (\vec{x} \cdot \vec{y})^2\\ &= (x_1y_1 + x_2y_2)^2\\ &= x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2\end{aligned}$$

Now we define

$$\begin{aligned}\phi(\vec{x}) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2) \in \mathbb{R}^3\\ \phi(\vec{y}) &= (y_1^2, \sqrt{2}y_1y_2, y_2^2) \in \mathbb{R}^3\end{aligned}$$

Then we have

$$\begin{aligned}\phi(\vec{x}) \cdot \phi(\vec{y}) &= x_1^2y_1^2 + (\sqrt{2}x_1x_2)(\sqrt{2}y_1y_2) + x_2^2y_2^2\\ &= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2\\ &= K(\vec{x}, \vec{y})\end{aligned}$$

This shows that $K(\vec{x}, \vec{y})$ is indeed a kernel function.

Example 2

Let

$$\begin{aligned}\vec{x} &= (x_1, x_2) \in \mathbb{R}^2\\ \vec{y} &= (y_1, y_2) \in \mathbb{R}^2\end{aligned}$$

We define

$$K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y} + \theta)^2.$$

We show that this is a kernel function. To do this, we note that

$$\begin{aligned}K(\vec{x}, \vec{y}) &= (\vec{x} \cdot \vec{y} + \theta)^2\\ &= (x_1y_1 + x_2y_2 + \theta)^2\\ &= \phi(\vec{x}) \cdot \phi(\vec{y})\end{aligned}$$

where

$$\phi(\vec{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2\theta}x_1, \sqrt{2\theta}x_2, \sqrt{\theta}) \in \mathbb{R}^6.$$

This shows that $K(\vec{x}, \vec{y})$ is indeed a kernel function.

10.10.3 Some important kernel functions

In the following we assume that $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$.

1. **Homogeneous polynomial kernel**

$$K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y})^d$$

   where d is some positive integer.

2. **Non-homogeneous polynomial kernel**

$$K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y} + \theta)^d$$

   where d is some positive integer and θ is a real constant.

3. **Radial basis function (RBF) kernel**

$$K(\vec{x}, \vec{y}) = e^{-\|\vec{x}-\vec{y}\|^2/2\sigma^2}$$

   This is also called the Gaussian radial function kernel.[1]

4. **Laplacian kernel function**

$$K(\vec{x}, \vec{y}) = e^{-\|\vec{x}-\vec{y}\|/\sigma}$$

5. **Hyperbolic tangent kernel function (Sigmoid kernel function)**

$$K(\vec{x}, \vec{y}) = \tanh(\alpha(\vec{x} \cdot \vec{y}) + c)$$

10.11 The kernel method (kernel trick)

10.11.1 Outline

1. Choose an appropriate kernel function $K(\vec{x}, \vec{y})$.
2. Formulate and solve the optimization problem obtained by replacing each inner product $\vec{x} \cdot \vec{y}$ by $K(\vec{x}, \vec{y})$ in the SVM optimization problem.
3. In the formulation of the classifier function for the SVM problem using the inner products of unclassified data $\vec{z}$ and input vectors $\vec{x}_i$, replace each inner product $\vec{z} \cdot \vec{x}_i$ with $K(\vec{z}, \vec{x}_i)$ to obtain the new classifier function.

10.11.2 Algorithm

Algorithm of the kernel method

Given a two-class linearly separable dataset of N points of the form

$$(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots, (\vec{x}_N, y_N),$$

where the y_i's are either +1 or 1 and appropriate kernel function $K(\vec{x}, \vec{y})$:

Step 1. Find $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ which maximizes

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j)$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i > 0 \text{ for } i = 1, 2, \ldots, N.$$

Step 2. Compute $\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$.

Step 3. Compute $b = \frac{1}{2}\left(\min_{i:y_i=+1} K(\vec{w}, \vec{x}_i) + \max_{i:y_i=-1} K(\vec{w}, \vec{x}_i)\right)$.

Step 4. The SVM classifier function is given by $f(\vec{z}) = \sum_{i=1}^{N} \alpha_i y_i K(\vec{x}_i, \vec{z}) + b$.

$\boxed{\min_{i:y_i=+1} K(\vec{w}, \vec{x}_i) + \max_{i:y_i=-1} K(\vec{w}, \vec{x}_i)).}$