Next: About this document ... **Up:** Support Vector Machines (SVM) **Previous:** 1-Norm Soft Margin

Kernel Mapping

The algorithm above converges only for linearly separable data. If the data set is not linearly separable, we can map the samples $\mathbf{x}$ into a feature space of higher dimensions:

$$\mathbf{x} \longrightarrow \phi(\mathbf{x})$$

in which the classes can be linearly separated. The decision function in the new space becomes:

$$f(\mathbf{x}) = \phi(\mathbf{x})^T\mathbf{w} + b = \sum_{j=1}^{m} \alpha_j y_j (\phi(\mathbf{x})^T \phi(\mathbf{x}_j)) + b$$

where

$$\mathbf{w} = \sum_{j=1}^{m} \alpha_j y_j \phi(\mathbf{x}_j)$$

and b are the parameters of the decision plane in the new space. As the vectors $\mathbf{x}_i$ appear only in inner products in both the decision function and the learning law, the mapping function $\phi(\mathbf{x})$ does not need to be explicitly specified. Instead, all we need is the inner product of the vectors in the new space. The function $\phi(\mathbf{x})$ is a kernel-induced *implicit* mapping.

Definition: A kernel is a function that takes two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ as arguments and returns the value of the inner product of their images $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$:

$$K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2)$$

As only the inner product of the two vectors in the new space is returned, the dimensionality of the new space is not important.

The learning algorithm in the kernel space can be obtained by replacing all inner products in the learning algorithm in the original space with the kernels:

$$f(\mathbf{x})=\phi(\mathbf{x})^T\mathbf{w}+b=\sum_{j=1}^m \alpha_j y_j K(\mathbf{x},\mathbf{x}_j)+b$$

The parameter b can be found from any support vectors $\mathbf{x}_i$:

$$b=y_i-\phi(\mathbf{x}_i)^T\mathbf{w}=y_i-\sum_{j=1}^m \alpha_j y_j(\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j))=y_i-\sum_{j=1}^m \alpha_j y_j K(\mathbf{x}_i,\mathbf{x}_j)$$

Example 0: linear kernel

Assume $\mathbf{x}=[x_1,\cdots,x_n]^T$, $\mathbf{z}=[z_1,\cdots,z_n]^T$,

$$K(\mathbf{x},\mathbf{z})=\mathbf{x}^T\mathbf{z}=\sum_{i=1}^n x_1 z_1$$

Example 1: polynomial kernels

Assume $\mathbf{x}=[x_1,x_2]^T$, $\mathbf{z}=[z_1,z_2]^T$,

$$\begin{aligned} K(\mathbf{x},\mathbf{z})=(\mathbf{x}^T\mathbf{z})^2 &: \quad (x_1z_1+x_2z_2)^2=x_1^2z_1^2+x_2^2z_2^2+2x_1z_1x_2z_2 \\ &: \quad <(x_1^2,x_2^2,\sqrt{2}x_1x_2),(z_1^2,z_2^2,\sqrt{2}z_1z_2)>=\phi(\mathbf{x})^T\phi(\mathbf{z}) \end{aligned}$$

This is a mapping from a 2-D space to a 3-D space. The order can be changed from 2 to general d.

Example 2:

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

Example 3:

$$K(\mathbf{x},\mathbf{z})=K(\mathbf{x},\mathbf{z})K(\mathbf{x},\mathbf{x})^{-1/2}K(\mathbf{z},\mathbf{z})^{-1/2}$$

Next: About this document ... **Up:** Support Vector Machines (SVM) **Previous:** 1-Norm Soft Margin

Ruye Wang 2016-08-19