Kernel Methods

Overview

- Kernel methods and the kernel function
- Dual representations
- Constructing kernels
- (Advanced topics):
  - Radial Basis Functions
  - Gaussian Processes

Kernel methods

These methods are based on the idea that training data can be used directly. They are sometimes called *memory based* methods.

- Training is fast
- Analysis/classification/prediction is very slow
- Based on the fact that there is a *metric* in the feature space, i.e. $||\mathbf{x}_i - \mathbf{x}_j||$ has meaning.

Kernel methods

Non-parametric methods (Parzen and kNN) use this approach. Classification can also be done directly with training data. For example, if we have a Bayesian estimation of parameters we get:

$$p(t|\mathbf{x}, \mathcal{D}) = \int_{\mathbf{w}} p(t|\mathbf{x}, \mathbf{w}) p(\mathbf{w}|\mathcal{D}) d\mathbf{w}$$

and our prediction/classification becomes

$$\hat{t} = E\{t|\mathbf{x}, \mathcal{D}\} = \int_{t} p(t|\mathbf{x}, \mathcal{D}) dt$$

no **w**

The kernel function

For models which are based on (most often nonlinear) feature space mappings $\phi(\mathbf{x})$the *kernel function* is given by:

$$k(\mathbf{x}, \mathbf{x}') = \boldsymbol{\phi}^T(\mathbf{x})\boldsymbol{\phi}(\mathbf{x}')$$

A very important property is that the kernel is *commutative* operation:

$$k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$$