Constructing Kernels

Constructing kernels

We need to be able to construct valid kernels if we are going to use the kernel trick. There are three approaches:

1. From basis functions
2. Guess and check
3. Use simpler kernel functions using rules

A necessary and sufficient condition for a kernel to be valid if its *Gram matrix* is positive semi-definite

Constructing kernels from basis functions

If we have defined basis functions for the data that we have we can simply use the formula

$$k(\mathbf{x}_n, \mathbf{x}_m) = \boldsymbol{\phi}_n^T \boldsymbol{\phi}_m$$

to construct the kernel function.

Construct kernels by guessing - example

For two dimensions we have:

$$
\begin{aligned}
k(\mathbf{x}, \mathbf{z}) =& (\mathbf{x}^T \mathbf{z})^2 \\
=& (x_1 z_1 + x_2 z_2)^2 \\
=& x_1^2 z_1^2 + 2 x_1 z_1 x_2 z_2 + x_2^2 z_2^2 \\
=& [x_1^2, \sqrt{2} x_1 x_2, x_2^2] \begin{bmatrix} z_1^2, \\ \sqrt{2} z_1 z_2 \\ z_2^2 \end{bmatrix} \\
=& \boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{z})
\end{aligned}
$$

so:

$$
\boldsymbol{\phi}(\mathbf{x}) = \begin{bmatrix} x_1^2, \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix}
$$

The basis function is 3 dimensional.

Constructing basis functions from simpler basis functions

We can use a set of rules (see eq. 6.13-22 in Bishop) to construct more complex kernels from simpler ones.

$\mathbf{x}^T\mathbf{z}$ is the simple linear kernel

$-\frac{1}{\sigma}\mathbf{x}^T\mathbf{z}$ is therefore a kernel because of 6.13

$\exp\{-\frac{1}{\sigma}\mathbf{x}^T\mathbf{z}\}$ is therefore a kernel because of 6.16

$$\exp(-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x})\exp(-\frac{1}{\sigma}\mathbf{x}^T\mathbf{z})\exp(-\frac{1}{2\sigma^2}\mathbf{z}^T\mathbf{z})$$

is a kernel because of 6.14

Constructing basis functions from simpler basis functions

We therefore get the Gaussian kernel:

$$
\begin{aligned}
k(\mathbf{x}, \mathbf{z}) &= \exp(-\frac{1}{2\sigma^2}||\mathbf{x} - \mathbf{z}||^2) \\
&= \exp(-\frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{z} + \mathbf{z}^T\mathbf{z})) \\
&= \exp(-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}) \exp(-\frac{1}{\sigma}\mathbf{x}^T\mathbf{z}) \exp(-\frac{1}{2\sigma^2}\mathbf{z}^T\mathbf{z}) \\
&= f(\mathbf{x}) k_0(\mathbf{x}, \mathbf{z}) f(\mathbf{z})
\end{aligned}
$$

The feature vector corresponding to the Gaussian kernel has *infinite dimensionality*.