

Expectation Maximization for Mixture Model

dw

September 18, 2021

1 Kernel Method

Kernel methods owe their name to the use of kernel functions, which enable them to operate in a high-dimensional, implicit feature space without ever computing the coordinates of the data in that space, but rather by simply computing the inner products between the images of all pairs of data in the feature space. This operation is often computationally cheaper than the explicit computation of the coordinates. This approach is called the "kernel trick".

For models which are based on a fixed nonlinear feature space mapping $\phi(\mathbf{x})$, the kernel function is given by the relation

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') \quad (1)$$

The simplest example of a kernel function is obtained by considering the identity mapping for the feature space in Eq. (1) so that $\phi(\mathbf{x}) = \mathbf{x}$, in which case $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$. We shall refer to this as the linear kernel.

For hyperplane (w, b) , the functional margin of sample point (x_i, y_i) is

$$\hat{\gamma}_i = y_i \left(\frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot x_i + \frac{b}{\|\mathbf{w}\|} \right) \quad (2)$$

$$\max_{\mathbf{w}, \mathbf{b}} \gamma \quad (3)$$

$$s.t. y_i (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) \geq \hat{\gamma} \quad (4)$$