

Figure 6.1 Illustration of the construction of kernel functions starting from a corresponding set of basis functions. In each column the lower plot shows the kernel function $k(x, x')$ defined by (6.10) plotted as a function of x for $x' = 0$, while the upper plot shows the corresponding basis functions given by polynomials (left column), 'Gaussians' (centre column), and logistic sigmoids (right column).

$$k(\mathbf{x}, \mathbf{x}') = ck_1(\mathbf{x}, \mathbf{x}') \quad (6.13)$$

$$k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') \quad (6.14)$$

$$k(\mathbf{x}, \mathbf{x}') = q(k_1(\mathbf{x}, \mathbf{x}')) \quad (6.15)$$

$$k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}')) \quad (6.16)$$

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \quad (6.17)$$

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}') \quad (6.18)$$

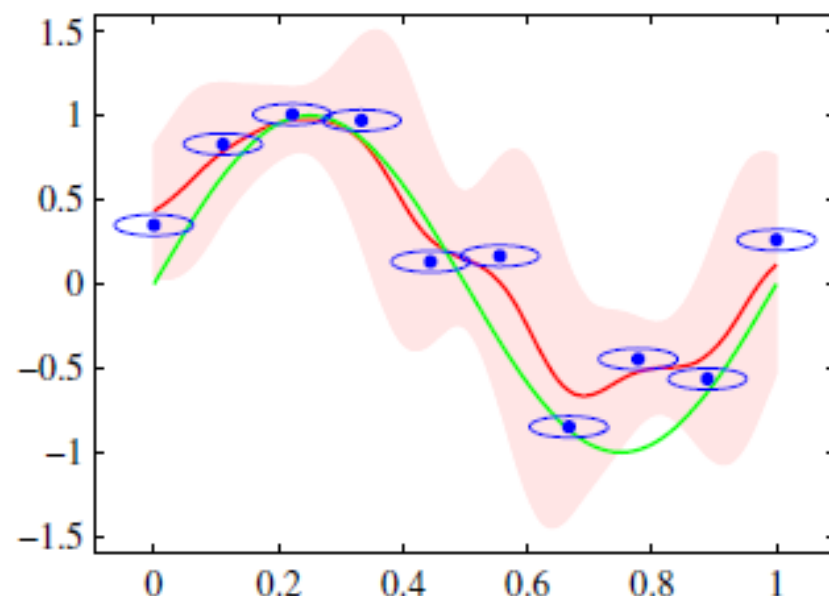
$$k(\mathbf{x}, \mathbf{x}') = k_3(\phi(\mathbf{x}), \phi(\mathbf{x}')) \quad (6.19)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{A} \mathbf{x}' \quad (6.20)$$

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a) + k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (6.21)$$

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a)k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (6.22)$$

Figure 6.3 Illustration of the Nadaraya-Watson kernel regression model using isotropic Gaussian kernels, for the sinusoidal data set. The original sine function is shown by the green curve, the data points are shown in blue, and each is the centre of an isotropic Gaussian kernel. The resulting regression function, given by the conditional mean, is shown by the red line, along with the two-standard-deviation region for the conditional distribution $p(t|x)$ shown by the red shading. The blue ellipse around each data point shows one standard deviation contour for the corresponding kernel. These appear noncircular due to the different scales on the horizontal and vertical axes.



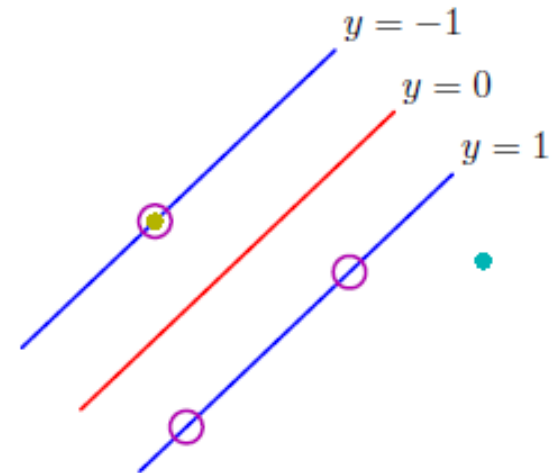
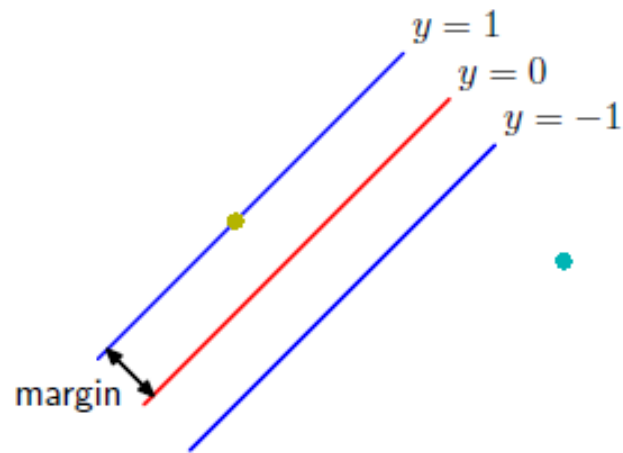
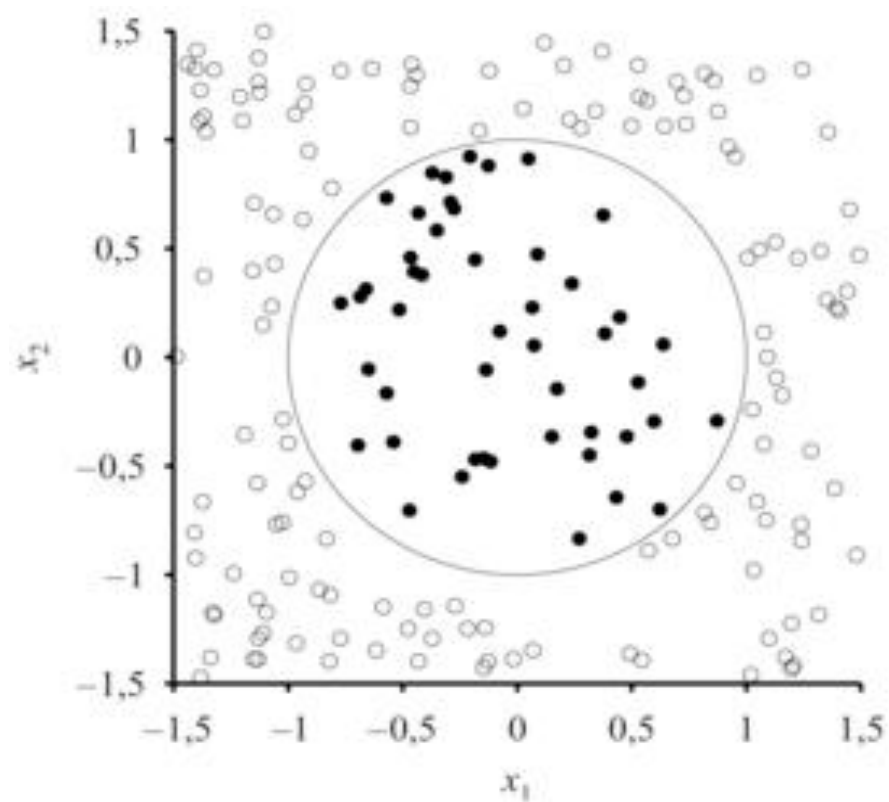
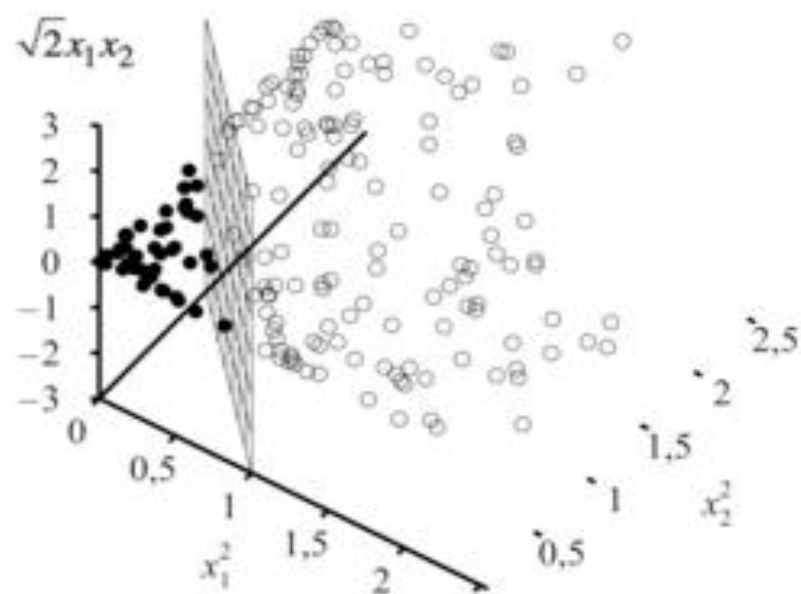


Figure 7.1 The margin is defined as the perpendicular distance between the decision boundary and the closest of the data points, as shown on the left figure. Maximizing the margin leads to a particular choice of decision boundary, as shown on the right. The location of this boundary is determined by a subset of the data points, known as support vectors, which are indicated by the circles.



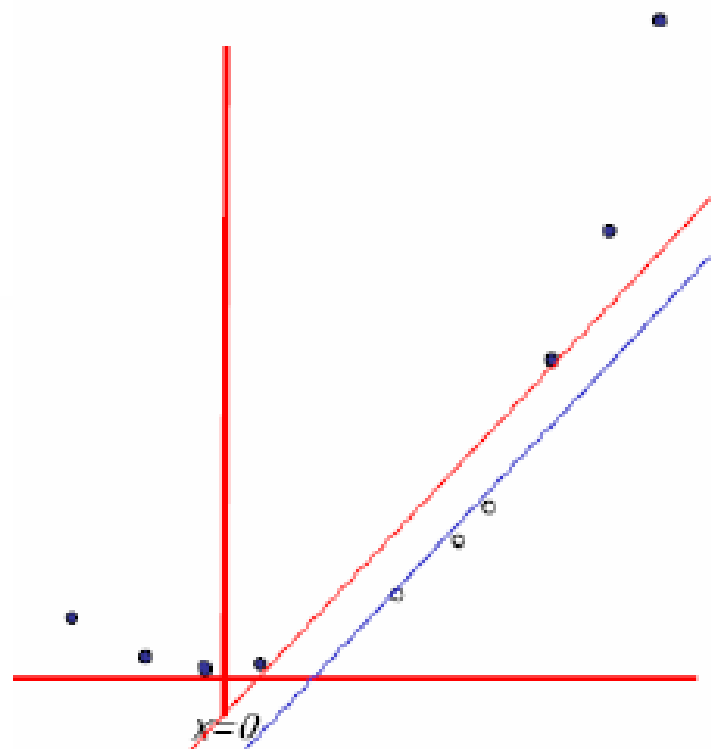
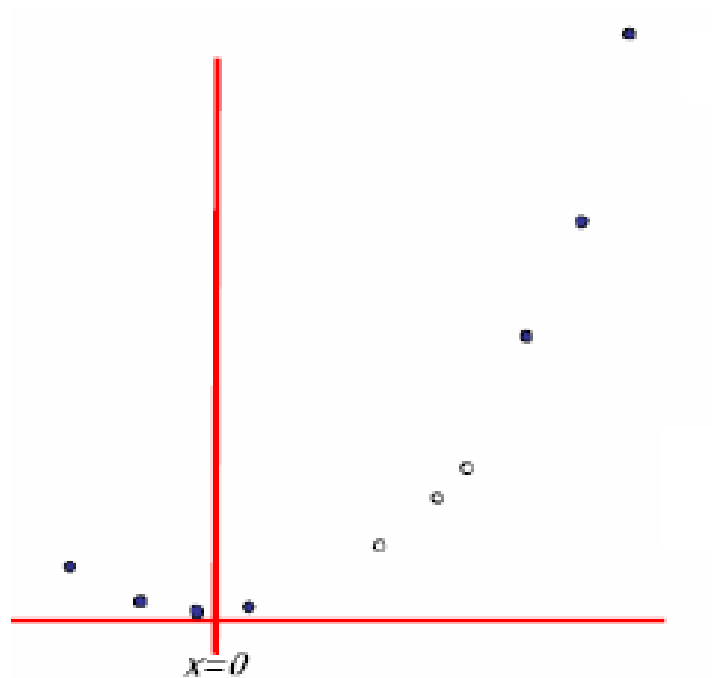
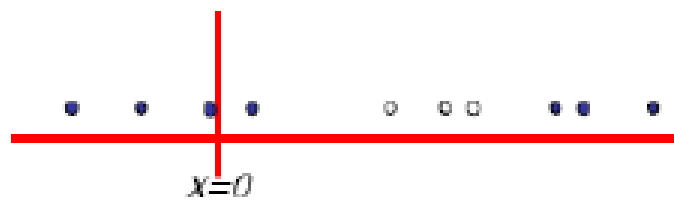
(a)



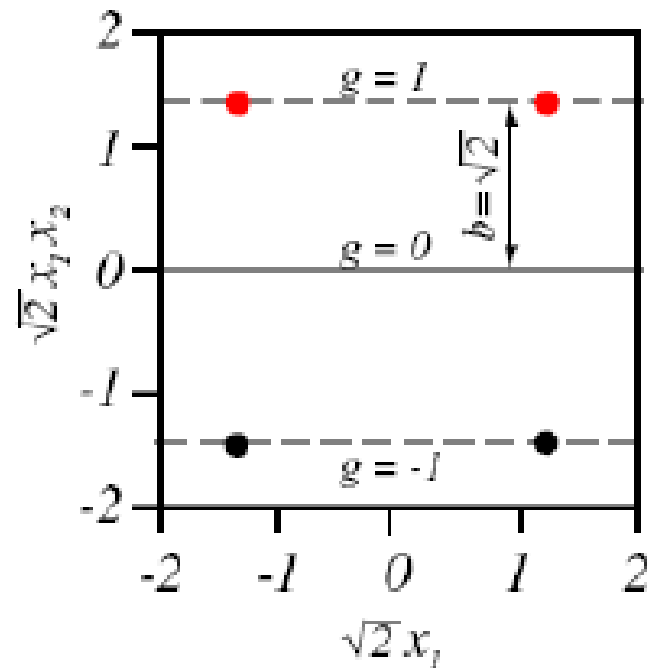
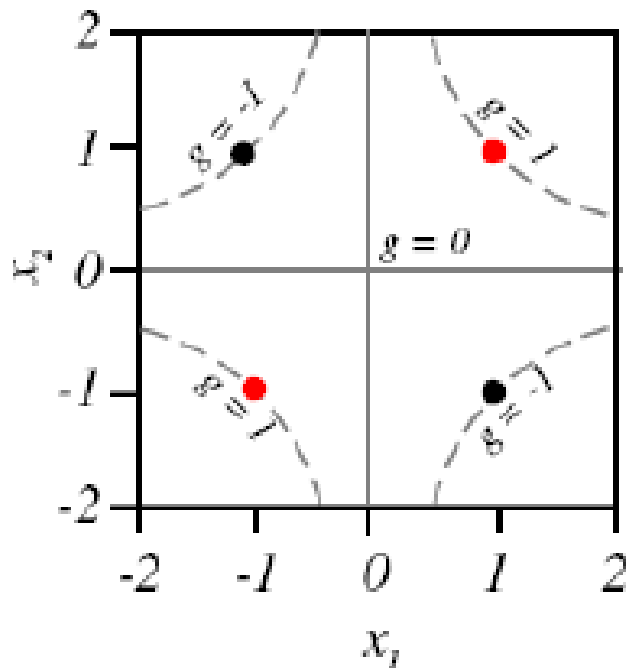
(b)

$$x_1^2 + x_2^2 \leq 1$$

$$x_1^2, x_2^2, \sqrt{2}x_1x_2$$



$$\mathbf{z}_k = (x_k, x_k^2)$$



$$\varphi: \mathbf{x} = (x_1, x_2) \rightarrow \varphi(\mathbf{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

$$\varphi(\mathbf{z}) \cdot \varphi(\mathbf{x}) = \dots = (1 + x_1z_1 + x_2z_2)^2 = (1 + \mathbf{z} \cdot \mathbf{x})^2 = k(\mathbf{z}, \mathbf{x})$$

$$\sqrt{2} x_1 x_2 = \pm \sqrt{2}$$

$$\varphi: \mathbf{x} = (x_1, x_2) \rightarrow \varphi(\mathbf{x}) = (\sqrt{2}x_1, \sqrt{2}x_1x_2)$$

$$x_2 = \pm \frac{1}{x_1}$$

Figure 7.2 Example of synthetic data from two classes in two dimensions showing contours of constant $y(x)$ obtained from a support vector machine having a Gaussian kernel function. Also shown are the decision boundary, the margin boundaries, and the support vectors.

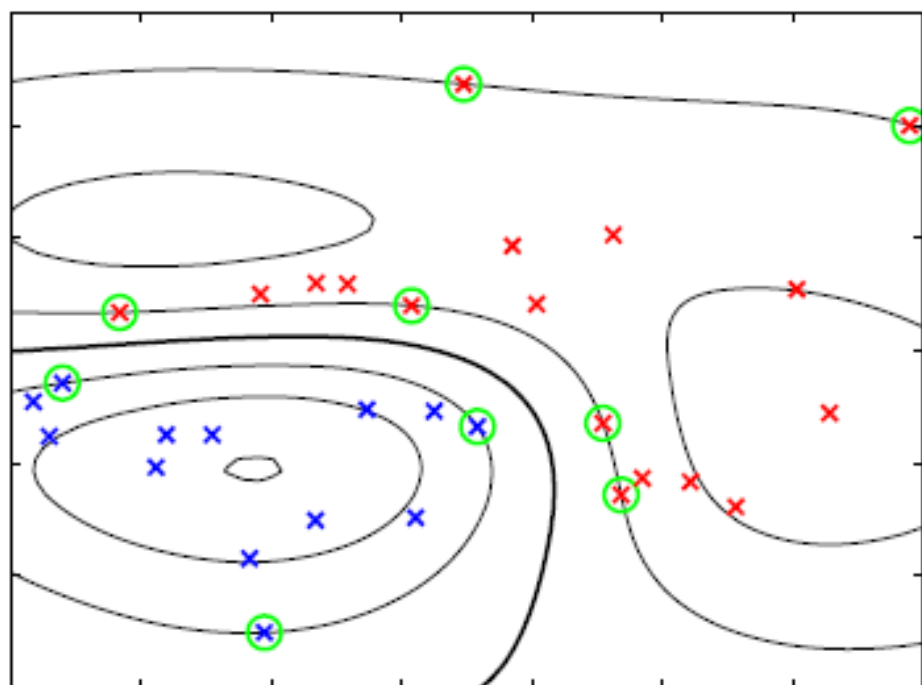


Figure 7.3 Illustration of the slack variables $\xi_n \geq 0$. Data points with circles around them are support vectors.

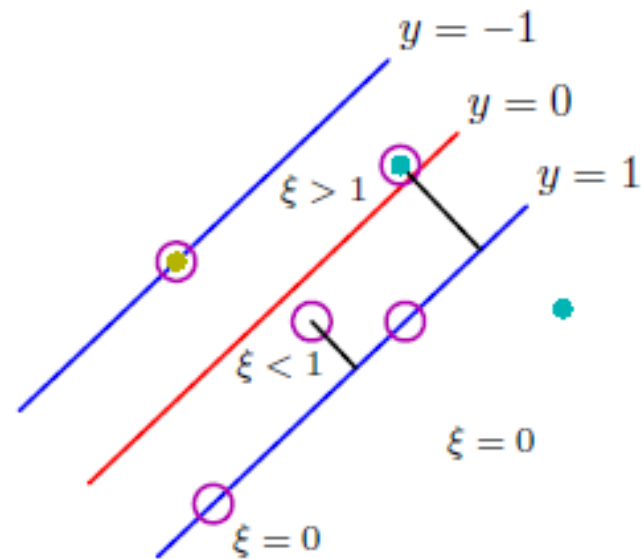


Figure 7.6 Plot of an ϵ -insensitive error function (in red) in which the error increases linearly with distance beyond the insensitive region. Also shown for comparison is the quadratic error function (in green).

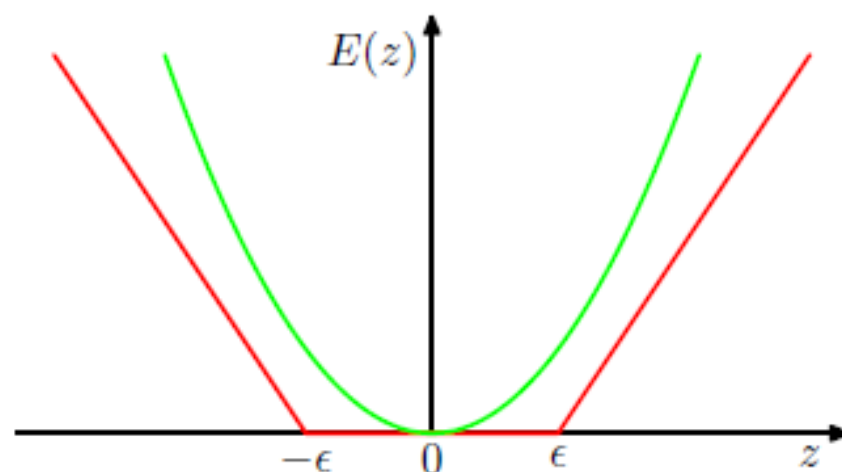


Figure 7.7 Illustration of SVM regression, showing the regression curve together with the ϵ -insensitive 'tube'. Also shown are examples of the slack variables ξ and $\hat{\xi}$. Points above the ϵ -tube have $\xi > 0$ and $\hat{\xi} = 0$, points below the ϵ -tube have $\xi = 0$ and $\hat{\xi} > 0$, and points inside the ϵ -tube have $\xi = \hat{\xi} = 0$.

