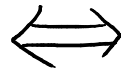


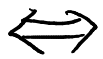
Support Vector Machine

Optimal Separating Hyperplane

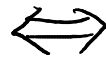
$$\begin{aligned} \max_{w, \beta, \|w\|=1} \quad & M \\ \text{s.t.} \quad & y_i(w^T x_i + \beta) \geq M \quad i=1, 2, \dots, N \end{aligned}$$



$$\begin{aligned} \max_{w, \beta} \quad & M \\ \text{s.t.} \quad & \frac{1}{\|w\|} (w^T x_i + \beta) \geq M, i=1, \dots, N \end{aligned}$$



$$\begin{aligned} \max_{w, \beta, \|w\|=\frac{1}{M}} \quad & M \\ \text{s.t.} \quad & y_i(w^T x_i + \beta) \geq 1 \quad i=1, 2, \dots, N \end{aligned}$$



$$\begin{aligned} \min_{w, \beta} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i(w^T x_i + \beta) \geq 1 \quad i=1, 2, \dots, N \end{aligned}$$

Dual Problem:

$$\min_{w, b} L(w, b, \alpha) = \min_{w, b} \frac{1}{2} \|w\|^2 + \sum_{i=1}^N \alpha_i (1 - y_i(w^T x_i + \beta))$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0 \quad w = \sum_{i=1}^N \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial \beta} = - \sum_{i=1}^N \alpha_i y_i = 0$$

$$(D^*): \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$\text{s.t.} \quad \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0$$

KKT Condition:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^N \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

$$y_i(w^T x_i + b) \geq 1$$

$$\alpha_i (y_i(w^T x_i + b) - 1) = 0$$

Slackened SVM.

Primal:

$$\begin{aligned} \min_{w, \beta} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0 \\ & y_i(w^T x_i + \beta) \geq 1 - \xi_i \end{aligned}$$

Dual:

$$L(w, b, \xi, \mu, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \alpha_i (y_i(w^T x_i + \beta) - 1 + \xi_i)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0 \quad w = \sum_{i=1}^N \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^N \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \mu_i - \alpha_i = 0 \quad \alpha_i = C - \mu_i$$

$$\text{then } \Theta_d(\mu, \alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$(D^*) \max_{\mu, \alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

$$\text{s.t. } \mu_i \geq 0, \alpha_i \geq 0$$

$$\alpha_i = C - \mu_i, \sum_{i=1}^N \alpha_i y_i = 0$$

KKT condition:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^N \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \mu_i - \alpha_i = 0$$

$$\alpha_i \geq 0, \mu_i \geq 0$$

$$\mu_i \xi_i = 0$$

$$\alpha_i (y_i(w^T x_i + \beta) - (1 - \xi_i)) = 0$$

$$1^\circ \xi_i > 0, \hat{\mu}_i = 0, \hat{\alpha}_i = C$$

$$2^\circ \hat{\alpha}_i < C, \hat{\mu}_i > 0, \xi_i = 0$$

SVM and kernels

$$\cdot \theta_D(\alpha, u) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle h(x_i), h(x_j) \rangle$$

$$\text{And } f(x) = w^T h(x) + \beta$$

$$= \sum_{i=1}^N \alpha_i y_i \langle h(x_i), h(x) \rangle + \beta$$

$$\cdot \text{Kernel: } K(x, x') = \langle h(x), h(x') \rangle$$

$$\text{d-th degree polynomials: } K(x, x') = (1 + \langle x, x' \rangle)^d$$

$$\text{Radial Basis: } K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

$$\text{Neural Network: } K(x, x') = \tanh(k_1 \langle x, x' \rangle + k_2)$$

