(a)

$$L_P = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i[y_i(x_i^T\beta+\beta_0)-(1-\xi_i)] - \sum_{i=1}^{N}\mu_i\xi_i, \quad (12.9)$$

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

$$\therefore \beta = \sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\frac{\partial L}{\partial \beta_0} = -\sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\therefore \sum_{i=1}^{N}\alpha_i y_i = 0$$

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

$$\therefore \alpha_i = C - \mu_i \quad , \ i=1,\dots,n$$

(b)

C (cost) : nonnegative tuning parameter

관측치들이 hyperplane 반대쪽으로 넘어가는 것에 얼만큼 penalty를 줄 것인가를 결정한다.

$$\Rightarrow \min_{\beta, \beta_0} \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N}\xi_i$$

subject to $\xi_i \geq 0$, $y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i \;\; \forall_i$

i) $C = 0 \Rightarrow \sum_{i=1}^{N}\xi_i = 0$

$\Rightarrow$ hyperplane은 당연하고, Margin을 넘어서는 것조차 허용하지 않는다.

= Maximal margin hyperplane optimization problem

ii) $C > 0$

① C가 작을 때

어찌됐든 $\frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N}\xi_i$ 를 최소화시키기 위해

Margin을 줄이는 역할을 한다.

(하나 이상의 양의 실수 ξ_i들의 크기를 줄여야 하므로)

⇒ highly fit to the data

⇒ low bias, high variance 과적합 발생 위험

② C가 클 때

반대로 Margin을 늘리는 역할을 한다.

⇒ less hard fit to the data

⇒ highly biased, lower variance

∴ C는 bias-variance tradeoff 기능을 담당한다.

(c)

$$0 \le \alpha_i \le C$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i [y_i(x_i^T\beta + \beta_0) - (1-\xi_i)] = 0,$$

$$\mu_i \xi_i = 0$$

$$y_i(x_i^T\beta + \beta_0) - (1-\xi_i) \ge 0$$

① $\alpha_i = 0 \Rightarrow \mu_i = C > 0 \Rightarrow \xi_i = 0$

$\Rightarrow y_i(x_i^T\beta + \beta_0) - (1-\xi_i) > 0$

$\Rightarrow y_i(x_i^T\beta + \beta_0) > 1 - \xi_i = 1$

해당 점은 Margin을 넘지 않고 경계 안쪽에 위치한다.

② $\alpha_i = C > 0 \Rightarrow \mu_i = 0 \Rightarrow \xi_i > 0$

$\Rightarrow y_i(x_i^T\beta + \beta_0) = 1 - \xi_i < 1$

해당 점은 Margin 넘어에 위치한다.

③ $0 < \alpha_i < C \Rightarrow \mu_i > 0 \Rightarrow \xi_i = 0$

$\Rightarrow y_i(x_i^T\beta + \beta_0) = 1 - \xi_i = 1$

해당 점은 Margin 선 상에 위치한다.