Generalization error of SVM

$$\min_{\{w,b\}} \quad \frac{1}{2}\|w\|^2 \qquad w^* = \sum_{i=1}^{N} \lambda_i y_i x^{(i)}$$

$$y_i(w^T x^{(i)} + b) \geq 1 \qquad \lambda_i \geq 0$$

$$i = 1, \ldots, N \qquad SV = \{i \mid \lambda_i > 0\}$$

Let $b = 0$

Optimum attained at w^*

$$\min_{\{w,b\}} \quad \frac{1}{2}\|w\|^2$$

$$y_i(w^T x^{(i)}) \geq 1 \qquad i = 1, \ldots N$$

$$\Rightarrow \quad \max_{w,\nu} \ \nu \qquad y_i(w^T x^{(i)}) \geq \nu \|w\| \quad i \in [N]$$

optimum (ν^*, w^*)

$$\nu^* = \frac{1}{\|w^*\|}$$

$$h_{\mathcal{D}}^{SVM}(x) = \operatorname{sign}(w^{*T} x)$$

$$R(h_{\mathcal{D}}^{SVM}) = P(\operatorname{sign}(w^{*T} X) \neq Y)$$

$$(X, Y) \sim P$$

A VC dimension approach

$$R(h) \leq R_{emp}(h) + \sqrt{\frac{V}{N}\left(\log N + \log \frac{1}{\delta}\right)}$$

with prob $1-\delta$.

$$R_{emp}(h) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{h(x^{(i)}) \neq y^{(i)}\}$$

Training set error

$V \rightarrow$ VC Dimension

[Reading : Burges Tutorial]

$H = \{h \mid h$ is a classifier$\}$

$V(H)$ = maximum number of points on which classifiers from H can learn all possible Labellings.

$V = d+1$

$$H = \left\{ h(x) = \text{sign}(w^T x + b) \;\middle|\; \begin{matrix} w \in \mathbb{R}^d \\ b \in \mathbb{R} \end{matrix} \right\}$$

If $\|w\| \leq B$. $\quad \|x\| \leq R \quad x \in \mathbb{R}^d$

$$H = \{ h \mid h(x) = \text{sign}(w^T x + b) \}$$

$$V(H) \leq B^2 R^2$$

$$R(h) \leq R_{emp}(h) + O\left(\frac{RB}{\sqrt{n}}\right)$$

with prob $1-\delta$.

$$\min_{\|w\| \leq B} R_{emp}(h)$$

$$h(x) = \text{sign}(w^T x + b)$$

Motivates the following formulation

$$\min_{w,b} \quad C \sum_{i=1}^{N} \max(0, 1 - y_i(w^T x^{(i)} + b)) + \frac{1}{2}\|w\|^2$$

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$y_i(w^T x^{(i)} + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

$$\mathcal{L}(w, b, \xi, \lambda, \mu) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \lambda_i \{y_i (w^T x^{(i)} + b) - 1 + \xi_i\} - \sum_{i=1}^{n} \mu_i \xi_i$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^{n} \lambda_i y_i x^{(i)} = 0$$

KKT

$$\frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^{n} \lambda_i y_i = 0$$

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

$$w = \sum_{i=1}^{n} \lambda_i y_i x^{(i)} \qquad C = \lambda_i + \mu_i$$

$$\sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\lambda_i \{ y_i (w^T x^{(i)} + b) - 1 + \xi_i \} = 0$$

$$y_i (w^T x^{(i)} + b) - 1 + \xi_i \geq 0$$

$$\xi_i \geq 0$$

$\lambda_i = 0 \Rightarrow \mu_i = C \Rightarrow \xi_i = 0$

$\Rightarrow y_i(w^T x^{(i)} + b) \geq 1$

$0 < \lambda_i < C \Rightarrow 0 < \mu_i < C \Rightarrow \xi_i = 0$

$\Rightarrow y_i(w^T x^{(i)} + b) = 1$

$\lambda_i = C \Rightarrow \mu_i = 0 \Rightarrow \xi_i \geqslant 0$

$\Rightarrow y_i(w^T x^{(i)} + b) - 1 + \xi_i = 0$

$\xi_i > 0 \Rightarrow \mu_i = 0 \Rightarrow \alpha_i = C = - - .$

$$\max_{w, b, \xi, \lambda, \mu} \mathcal{L}(w, b, \xi, \lambda, \mu)$$

$$w = \sum_{i=1}^{n} \lambda_i y_i x^{(i)}, \quad \sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\lambda_i + \mu_i = C$$

$$\lambda_i \geqslant 0, \; \mu_i \geqslant 0 \quad i = 1, \ldots, n$$

$$\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i y_i x^{(i)T} \omega - \cdot$$

$$= -\frac{1}{2}\left\|\sum_{i=1}^{n} \lambda_i y_i x^{(i)}\right\|^2$$

$$= -\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x^{(i)T} x^{(j)} \cdot$$

$$\max_{\lambda \geq 0} \sum_i \lambda_i - \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j x^{(i)T} x^{(j)}$$

$$0 \leq \lambda_i \leq C, \quad \sum_i \lambda_i y_i = 0$$