Lecture 7

Chiranjib Bhattacharyya

chibha@chalmers.se

February 14, 2012

Agenda

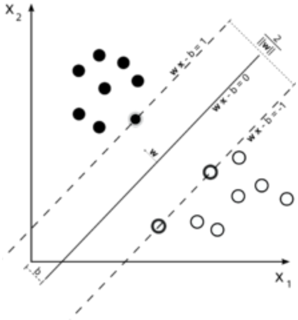
- Support vector machines

Support Vector Machines

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

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

SVM



Recall

- The class $y = 1$ lies in the halfspace

$$w^\top x + b \geq 1$$

- The class $y = -1$ lies in the halfspace

$$w^\top x + b \leq -1$$

- The support vectors either lie in $w^\top x + b = 1$ or $w^\top x + b = -1$

SVMs(Non-Separable case)

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

$$y_i(w^\top x_i + b) \geq 1 - \xi_i \;\; \xi_i \geq 0 \; i = 1, \ldots, n$$

SVMs(Non-separable case)

At optimality $\alpha_i \geq 0$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$b = y_i - w^\top x_i \forall C > \alpha_i > 0$$

$$\xi_i > 0 \Leftarrow \alpha_i = C$$

SVM Dual

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j$$

$$C \geq \alpha_i \geq 0 \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Non-Linear Classifiers

Let $x = [x_1, x_2]^\top$.
Define $\Phi(x) = [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2]^\top$
See that $\Phi(u)^\top \Phi(v) = (1 + u_1 v_1 + u_2 v_2)^2$
Computing a linear SVM on $\Phi(x)$.

Non-Linear Classifiers

SVM Dual

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$C \geq \alpha_i \geq 0 \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$f(x) = \text{sign}(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b)$$

Some facts

Let K_1, K_2 be two kernels

$$K(u,v) = \eta_1 K_1(u,v) + \eta_2 K_2(u,v) \;\; \eta_1, \eta_2 \geq 0$$

$$K(u,v) = K_1(u,v) K_2(u,v)$$

$$K(u,v) = \frac{K_1(u,v)}{\sqrt{K_1(u,u)}\sqrt{K_1(v,v)}}$$

Some facts

Let K_1, K_2 be two kernels

$$K(u,v) = \eta_1 K_1(u,v) + \eta_2 K_2(u,v) \;\; \eta_1, \eta_2 \geq 0$$

$$K(u,v) = K_1(u,v) K_2(u,v)$$

$$K(u,v) = \frac{K_1(u,v)}{\sqrt{K_1(u,u)}\sqrt{K_1(v,v)}}$$

Popular Kernels

$$K(u,v) = (1 + u^\top v)^p \text{ polynomial kernel of degree} p$$

$$K(u,v) = e^{-\frac{\|u-v\|^2}{\sigma^2}} \text{ radial basis kernel}$$

Connection to Generalization error

$\mathscr{F}$ be the function class.
f be a classifier

$$R_n(f) = \frac{1}{n}\sum_{i=1}^{n} I_{f(x_i) \neq y_i}$$

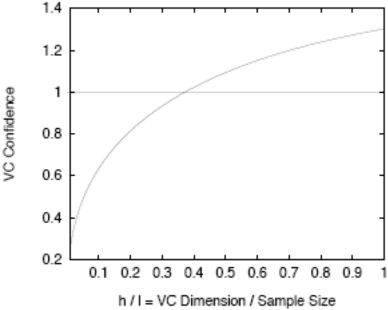
$$R(f) = P(f(x) \neq Y)$$

Generalization error

$$R(f) \leq R_n(f) + g(h(\mathscr{F}), \delta, n)$$

with probability atleast $1 - \delta$

Connection to Generalization error



Prediction problems

Data: $D = \{(x_i, y_i) | i = 1, \ldots, n\}$
Find w, b such that

$$y_i = w^\top x_i + b$$

Regression formulation

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

$$-\varepsilon \leq y_i - w^\top x_i - b \leq \varepsilon$$

SVM Regression formulation

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

$$y_i - w^\top x_i - b \leq \varepsilon + \xi_i$$

$$w^\top x_i + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i \geq 0, \xi_i^* \geq 0$$

The Dual formulation

$$max_{\alpha,\alpha^*} \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*) x_i^\top x_j$$

$$\sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \alpha^* \;\; 0 \leq \alpha_i, \alpha_i^* \leq C$$

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) x_i$$

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) x_i^\top x + b$$