SVM

Originally: Maximize the margin

$$\max_{\vec{w},b} \frac{1}{\|w\|}$$

$$\text{s.t. } 1 - y_i(\vec{x}_i \cdot \vec{w} + b) \leq 0$$

Primal version by applying Lagrange multiplier:

$$\min_{\vec{w}} \max_{\vec{\alpha}} \|\vec{w}\|/2 + \sum_i^n \alpha_i \left(1 - y_i(\vec{x}_i \cdot \vec{w} + b)\right)$$

$$\text{s.t. } \alpha_i \geq 0 \ \forall i$$

Dual version (swap max/min + apply KKT condition)

$$\max_{\vec{\alpha}} \sum_i^n \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$$

$$\text{s.t. } \begin{cases} \alpha_i \geq 0 & \forall i \\ \sum_i^n \alpha_i y_i = 0 \end{cases}$$

KKT conditions: of (w^*, α^*)

$$\begin{cases} \frac{\partial L(w^*, \alpha^*)}{\partial w_i} = 0 & \forall i \leq d \\ \alpha_i^* g_i(w^*) = 0 & \forall i \leq n \\ g_i(w^*) \leq 0 & \forall i \leq n \\ \alpha^* \geq 0 & \forall i \leq n \end{cases}$$

$\left(g_i(\vec{w}) = 1 - y_i(\vec{w}\vec{x}_i + b)\right)$

Slack penalty (soft-SVM)

$$\underset{\vec{w}, \vec{\varepsilon}, b}{\text{minimize}} \ \frac{\|w\|^2}{2} + C\sum \varepsilon_i$$

$$\text{s.t.} \begin{cases} (\vec{w}\vec{x}_i + b)\, y_i \geq 1 - \varepsilon_i & \forall i \\ \varepsilon_i \geq 0 \end{cases}$$

Primal (Lagrangian) form

$$\underset{\vec{w}, \vec{\varepsilon}, b}{\text{minize}} \ \max_{\vec{\alpha}} \frac{1}{2}\|w\|^2 + C\sum \varepsilon_i$$

$$+ \sum \alpha_i \left[1 - \varepsilon_1 - y_i(\vec{w}\vec{x}_i + b)\right]$$

$$\text{s.t.} \begin{cases} \alpha_i \geq 0 & \forall i \\ \varepsilon_i \geq 0 & \forall i \end{cases}$$

Dual form

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

$$\text{s.t.} \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$$

Bayes net

① FOD

Learning: $\theta^*_{x_i|pa(x_i)} = \frac{\#(x_i, pa(x_i))}{\#(pa(x_i))}$ sufficient statistic

② POD : EM algorithm.

E-step: softly assign values to missing vars

M-step: update model's params based on soft labels

(Initially, randomly initialize)

Loss function

Hinge loss $= (1-yf)_+ = \max(0, 1-yf)$

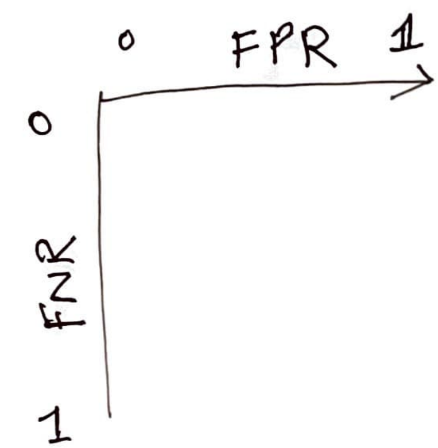
Binomial deviance $= \log(1+\exp(-2yf))$

☆ SVM Soft-margin loss $\equiv \min_{\vec{w}, b} \frac{1}{2}\|w\|^2$ ~~+ C~~ L2-regularizer

$+ C \sum_{i=1}^{n} \max(0, 1 - y_i(\vec{w}^T x_i + b))$ hinge-loss

		Prediction	
		1	0
Actual	1	TP	FN
	0	FP	TN

$$F1 = \frac{2 \times P \times R}{P+R}$$



$$FPR = \frac{FP}{FP+TN} \rightarrow \text{actual } N$$

$$FNR = \frac{FN}{FN+TP} \rightarrow \text{actual } P$$

$$TPR = \frac{TP}{TP+FN} \rightarrow \text{actual } P$$

(i.e., Sensitivity, Recall)

$$TNR = \frac{TN}{TN+FP} \rightarrow \text{actual } N$$

(i.e, Specificity)