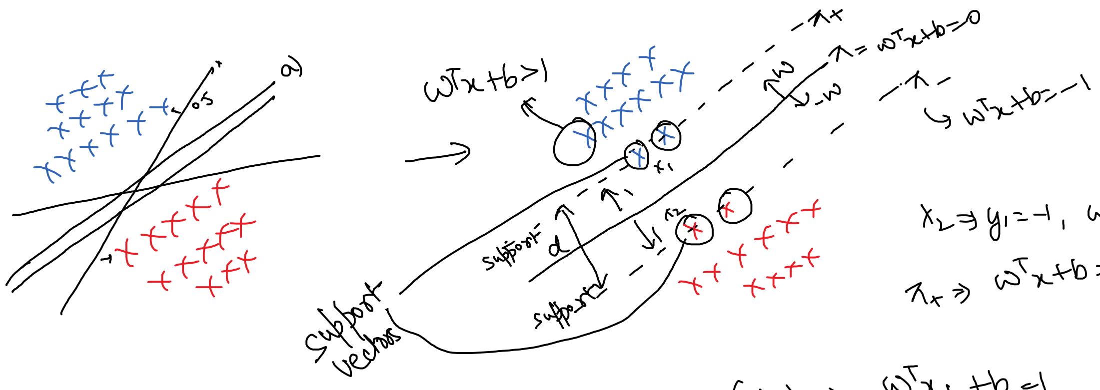
Saturday, May 25, 2024 9:35 AM

Support Vector Machine



$x_2 \Rightarrow y_1=-1,\ w^Tx_1+b=-1$

$\pi_+ \Rightarrow w^Tx+b=1$

$\text{pred}\ (x_1) \Rightarrow w^Tx_1+b=1$

$w^Tx>0$ (Blue) , $w^Tx<0$ (Red)

$$\pi^+ :\ w^Tx_1+b=1$$

$$\pi^- :\ \underset{-}{w^Tx_2}\ \underset{-}{+b}\ =\ \underset{+}{-1}$$

$$w^T(x_1-x_2) = 2$$

Normalizing the above eqⁿ

$$\frac{w^T}{||w||}(x_1-x_2) = \frac{2}{||w||}$$

$$\boxed{x_1-x_2 = \frac{2}{||w||} = d} \rightarrow \text{Margin}$$

Mathematical Objective Fⁿ:

$$w^*, b^* = \underset{w,b}{\text{argmax}}\left(\frac{2}{||w||}\right)$$

Building Constraints:

(I) Blue in its zone

$$y_i(w^Tx_i+b)=1$$

(II) Red in its zo

$$y_i(w^Tx_i+b)=1$$

① Blue in its zone

$y_i(w^T x_i + b) = 1$

$\hookrightarrow y_i = 1$, $w^T x + b = 1$

⑪ $y_i(w^T x_i + b) = 1$

$\downarrow -1$ $\downarrow -1$

(III) $y_i(w^T x + b) > 1$

(IV) $y_i(w^T x_i + b) > 1$

$\hookrightarrow -1$ $\hookrightarrow <1$

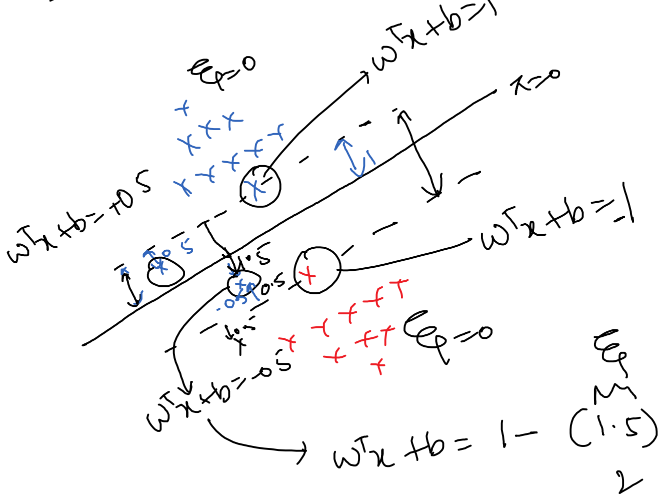
MoF

$$w^*, b^* = \underset{w,b}{\text{argmax}} \frac{2}{||w||} \text{ such that}$$

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

The above is hard margin.

Reality



$\xi \Rightarrow$ represents misclassification

$\downarrow$

distance of a datapoints from its correct hyperplane in its opp direction.

MoF

$$w^*, b^* = \underbrace{\text{argmax} \left(\frac{2}{||w||}\right)}_{\text{regularizer}} + \overset{\overset{\text{hyperparameter}}{\uparrow}}{C} \underline{\sum \xi_i} \text{ such that}$$

loss

$$\overbrace{y_i(w^T x_i + b)}^{z_i} \geq 1 - \xi_i$$

$$w^*, b^* = \underset{w}{\operatorname{argmin}} \underbrace{\left(\frac{\|w\|}{2}\right)}_{\text{Regularizer}} + C \underbrace{\sum \xi_i}_{\text{loss}}$$

logistic = loss f^n + λ regularizer

SVM = C loss f^n + regularizer

$$C \propto \frac{1}{\lambda}$$

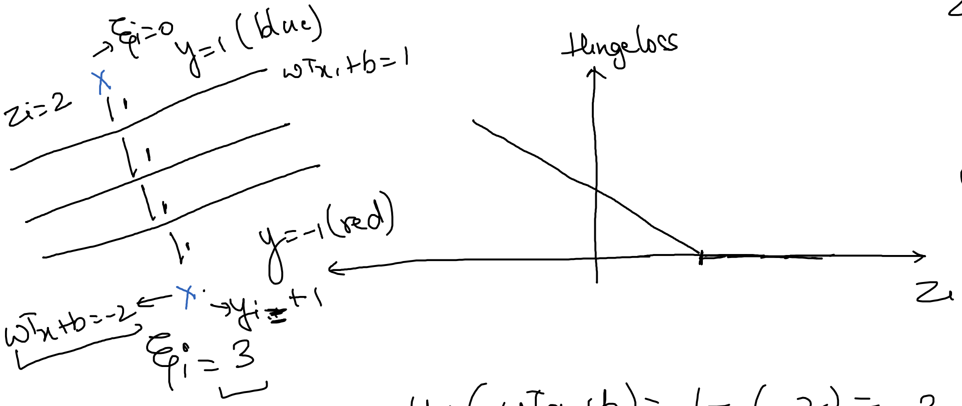
$$\left.\begin{matrix} lR \\ lgR \end{matrix}\right] \rightarrow C$$

1) C↓ → more error → underfitting

2) C↑ → less error → overfitting

Loss Minimization: (Hingeloss) ⇒ $\max(0, 1 - z_i)$

$$z_i = y_i(w^T x_i + b)$$



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

$$y_i(w^T x + b) = 1 - (\underbrace{-2}) = \underbrace{3} = \xi_i$$

$$\boxed{\xi_i = 1 - z_i}$$

Hingeloss = $\max(0, \xi_i)$

for the misclassified point,

Loss = max(0, 3) = 3

for correctly classified point,

Loss = $\max(0, 1 - z_i)$

$= \max(0, 1-2) = \max(0, -1)$

$= 0$

Dual form of SVM:

Primal form: $w^*\ b^* = \underset{w}{\operatorname{argmin}} \left(\frac{||w||}{2} \right) + C \sum \xi_i$

Dual form: $\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \underbrace{(x_i^T x_j)}_{\text{Kernel trick}} \xrightarrow{} \text{similarity}$

$x_i \rightarrow \alpha_i$

$\alpha_i = 0$ for all pts except support vectors

for supports $\alpha_i > 0$

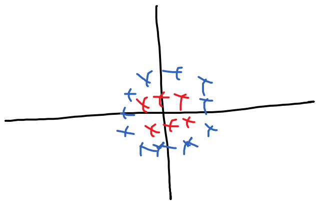
Kernels

↳ Mathematical functions

linear

polynomial

RBF
Radial Basis Function



Kernel trick: transform your datapoints by applying functions to make linearly seperable

Polynomial Kernel: $K(x_1, x_2) = (x_1^T x_2 + c)^d$

$d \rightarrow$ degrees

$d=2 \rightarrow$ Quadratic Eq^n ax^2+bx+c

$X_1 = [x_{11} \quad x_{12}]$ $\quad X_2 = [x_{21} \quad x_{22}]$

$c=1, d=2$

$$(1 + x_1^T x_2)^2 \Rightarrow \left(1 + [x_{11}\ x_{12}] \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}\right)^2$$

$$\Rightarrow (\overset{a}{1} + \overset{b}{x_{11}x_{21}} + \overset{c}{x_{12}x_{22}})^2$$

$\hookrightarrow (a+b+c)^2$

$= a^2+b^2+c^2+2ab+2bc+$
$2ca$

$$\Rightarrow (1^2 + \underline{x_{11}^2}x_{21}^2 + \underline{x_{12}^2}x_{22}^2 + \overset{\sqrt{2}\times\sqrt{2}}{2}x_{11}x_{21} + \underline{2x_{11}}x_{21}\underline{x_{12}}x_{22} + \underset{\sqrt{2}\sqrt{2}}{2}x_{12}x_{22})$$

$[1, x_{11}^2, x_{12}^2, \sqrt{2}x_{11}, \sqrt{2}x_{11}x_{12}, \sqrt{2}x_{12}]$

$[1, x_{21}, x_{22}, \sqrt{2}x_{21}, \sqrt{2}x_{21}x_{22}, \sqrt{2}x_{22}]$

$x_1' \rightarrow 6$ axis
$\hookrightarrow 6d \leftarrow$ transform

$X_2' \rightarrow 6$ axis
$\hookrightarrow 6d$

$\hookrightarrow 6d \xleftarrow{\text{transform}}$ $\quad\quad\quad\quad$ $\hookrightarrow 6d$

$x_1 \rightarrow [x_{11} \quad x_{12}] \rightarrow 2d$

** **Mercer's Theorem:** Kernel converts the d-dim dataset into d'-dim dataset such that d' > d

RBF Kernel: $K(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}}$

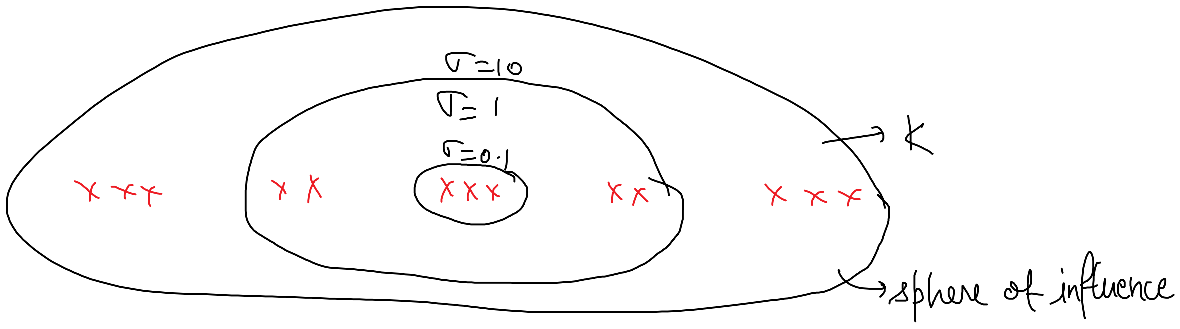
(I) $\|x_1 - x_2\| = d$

$$K = e^{-d^2/2\sigma^2}$$

$$d\uparrow \rightarrow d^2 \rightarrow \frac{d^2}{2\sigma^2}\uparrow \rightarrow \frac{1}{e^{d^2/2\sigma^2}\uparrow}\downarrow \rightarrow K\downarrow$$

(II) $K = \frac{1}{e^{d^2/2\sigma^2}}$

$$\sigma\uparrow \rightarrow \sigma^2\uparrow \rightarrow \frac{d^2}{2\sigma^2}\downarrow \rightarrow e^{d^2/2\sigma^2}\downarrow \rightarrow K\uparrow$$



Hyperparameter ⇒ gamma → $\frac{1}{2\sigma^2}$ ⇒ $e^{d\gamma}$

$\sigma\uparrow \rightarrow \sigma^2\uparrow \rightarrow \gamma\downarrow \rightarrow e^{d\gamma}\downarrow \rightarrow \frac{1}{e^{d\gamma}}\uparrow \rightarrow K\uparrow$