MULTICLASS KERNEL MACHINES

$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ $y_i \in \{1, 2, \ldots, K\}$

one-versus-all $\Rightarrow$

training set size $= N$

$\left.\begin{array}{l} 1^{+} \text{ vs } \{2, \ldots, K\}^{-} \\ 2^{+} \text{ vs } \{1, 3, \ldots, K\}^{-} \\ \vdots \\ K^{+} \text{ vs } \{1, 2, \ldots, K-1\}^{-} \end{array}\right]$ $\begin{array}{l} \rightarrow svm_1 \\ \rightarrow svm_2 \\ \vdots \\ \rightarrow svm_K \end{array}$

x^* $f_1(x^*)$ $f_2(x^*)$ $\ldots$ $f_K(x^*)$

pick the maximum

one-versus-other

$N_1 + N_2 < N$

$\left.\begin{array}{ll} f_{12} & 1 \text{ vs } 2 \\ f_{13} & 1 \text{ vs } 3 \\ & \vdots \\ f_{1K} & 1 \text{ vs } K \\ & 2 \text{ vs } 3 \\ f_{23} & 2 \text{ vs } 4 \\ f_{24} & \vdots \\ & 2 \text{ vs } K \\ \vdots & \vdots \\ f_{(K-1)K} & (K-1) \text{ vs } K \end{array}\right]$

$N_1 = \frac{N}{K}$

$N_2 = \frac{N}{K}$

$\left[\frac{2N}{K}\right]$ data points

$\frac{K(K-1)}{2}$

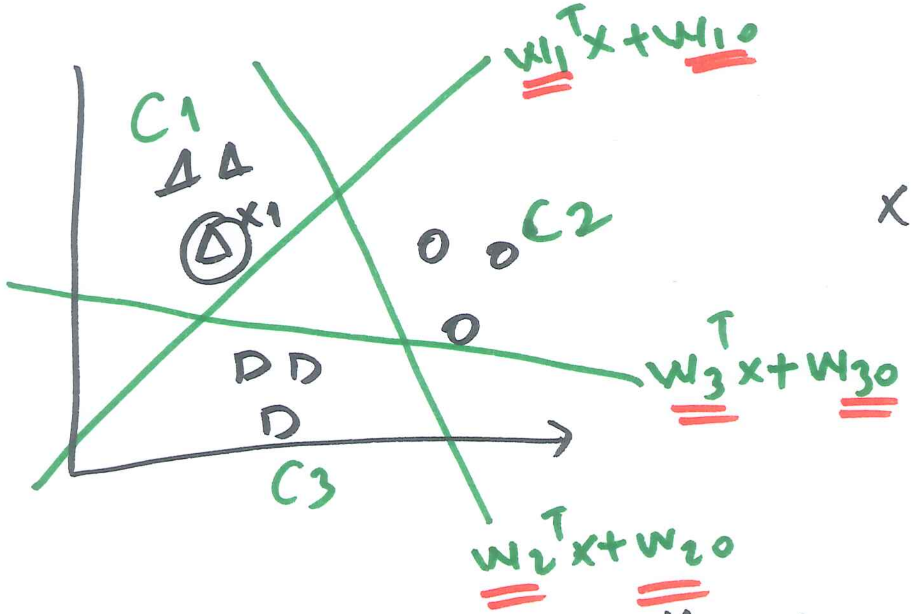
for example: 100 class classification problem

N data points $\otimes$ K classifiers.

$\frac{N}{50}$ data points. $\otimes$ $\frac{K(K-1)}{2}$ classifiers

$x^* \Rightarrow$ pick the one with highest # of wins.

$\underset{C\#1}{\underline{5}}$ $\underset{C\#2}{\underline{3}}$ — . . . — $\underset{C\#K}{\underline{2}}$



$$x_1 \Rightarrow \begin{cases} w_1^T x_1 + w_{10} \geq w_2^T . x_1 + w_{20} + 2 - \varepsilon_{12} \\ w_1^T x_1 + w_{10} \geq w_3^T . x_1 + w_{30} + 2 - \varepsilon_{13} \end{cases}$$

minimize $\quad \frac{1}{2} \sum_{c=1}^{K} \|w_c\|^2 + C \sum_{i=1}^{N} \sum_{c=1}^{K} \varepsilon_{ic}$

subject to: $\underset{N \times (K-1)}{\Big[} \; w_{y_i}^T x_i + w_{y_i 0} \geq w_c^T x_i + w_{c0} + 2 - \varepsilon_{ic} \quad \forall (i, c \neq y_i)$

$$\varepsilon_{ic} \geq 0 \quad \forall (i, c \neq y_i)$$

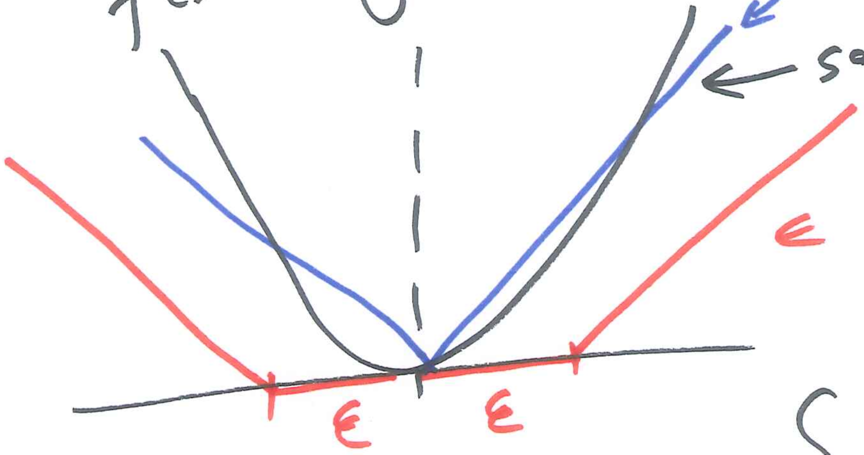
\# of decision variables :

$$= K(D+1) + N(K-1)$$

KERNEL MACHINES FOR REGRESSION

$f(x) = \hat{y} = w^T x + w_0$

squared error $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$



$\left.\begin{array}{l}\epsilon\text{-insensitive}\\ \text{loss}\end{array}\right\}$ if $\epsilon = 0 \Rightarrow$ absolute error.

$$\epsilon\text{-insensitive loss} = \begin{cases} 0 & \text{if } |y_i - \hat{y}_i| < \epsilon \\ \underbrace{|y_i - \hat{y}_i|}_{} - \underset{\sim}{\epsilon} & \text{if } |y_i - \hat{y}_i| \geq \epsilon \end{cases}$$

minimize $\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} (\varepsilon_i^+ + \varepsilon_i^-)$

subject to: $\alpha_i^+ \left[\, y_i - \overbrace{(w^T x_i + w_0)} \leq \epsilon + \varepsilon_i^+ \quad \cancel{\forall i}\right.$

$\alpha_i^- \left[\, (w^T x_i + w_0) - y_i \leq \epsilon + \varepsilon_i^- \quad \forall i\right.$

$\beta_i^+ \left[\, \varepsilon_i^+ \geq 0 \quad \forall i\right.$

$\beta_i^- \left[\, \varepsilon_i^- \geq 0 \quad \forall i\right.$

$$\boxed{\begin{array}{c} y_i \leq \hat{y}_i + \epsilon \\ \hline y_i - \epsilon \leq \hat{y}_i \end{array}}$$

$$L_p = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\varepsilon_i^+ + \varepsilon_i^-) - \sum_{i=1}^{N}\alpha_i^+\left[-y_i + (w^T x_i + w_0) + \epsilon + \varepsilon_i^+\right]$$

$$- \sum_{i=1}^{N}\alpha_i^-\left[-(w^T x_i + w_0) + y_i + \epsilon + \varepsilon_i^-\right] - \sum_{i=1}^{N}\beta_i^+ \cdot \varepsilon_i^+ - \sum_{i=1}^{N}\beta_i^-\varepsilon_i^-$$

$$\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{N}\alpha_i^+ x_i + \sum_{i=1}^{N}\alpha_i^- x_i = 0 \Rightarrow w = \sum_{i=1}^{N}\underline{(\alpha_i^+ - \alpha_i^-)}\, x_i$$

$\alpha_i . y_i$

$\sum \alpha_i y_i = 0$

$$\frac{\partial L_p}{\partial w_0} = -\sum_{i=1}^{N}\alpha_i^+ + \sum_{i=1}^{N}\alpha_i^- = 0 \Rightarrow \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-) = 0$$

$$\frac{\partial L_p}{\partial \varepsilon_i^+} = C - \alpha_i^+ - \beta_i^+ = 0 \Rightarrow \alpha_i^+ + \beta_i^+ = C \Rightarrow \begin{array}{l} 0 \le \alpha_i^+ \le C \\ 0 \le \beta_i^+ \le C \end{array}$$

$$\frac{\partial L_p}{\partial \varepsilon_i^-} = C - \alpha_i^- - \beta_i^- = 0 \Rightarrow \underbrace{\alpha_i^-}_{\ge 0} + \underbrace{\beta_i^-}_{\ge 0} = C \Rightarrow \begin{array}{l} 0 \le \alpha_i^- \le \epsilon \\ 0 \le \beta_i^- \le C \end{array}$$

Exercise

maximize $\sum_{i=1}^{N} y_i(\alpha_i^+ - \alpha_i^-) - \epsilon\sum_{i=1}^{N}(\alpha_i^+ + \alpha_i^-) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\boxed{x_i^T x_j}$

$x_i^T x_j \rightarrow K(x_i, x_j)$

subject to: $\sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-) = 0$

$0 \le \alpha_i^+ \le C \quad \forall i$

$0 \le \alpha_i^- \le C \quad \forall i$

# of decision variables: $2*N$

x^*

$$f(x^*) = w^T . x^* + w_0$$

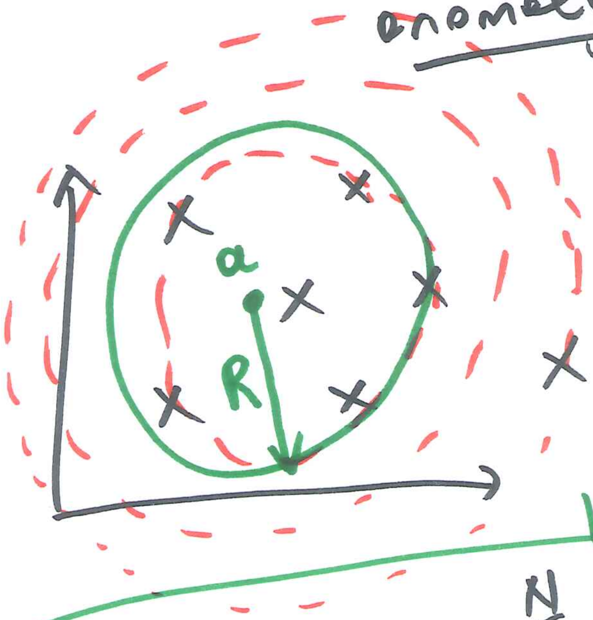
$$\boxed{= \sum_{i=1}^{N}(\alpha_i^+ - \alpha_i^-)\, x_i^T . x^* + w_0}$$

decision function.

ONE-CLASS KERNEL MACHINES

$\mathcal{X} = \{(x_i)\}_{i=1}^{N}$ test data points $\Rightarrow x \in \mathcal{X}$ or $x \notin \mathcal{X}$

anomaly detection or outlier detection



minimize $\underline{\underline{R^2}} + C \sum_{i=1}^{N} \varepsilon_i$

subject to: $\|x_i - a\|^2 \leq R^2 + \varepsilon_i \;\; \forall i$

$\varepsilon_i \geq 0 \;\; \forall i$

Exercise

$a = \ldots\ldots$

$R = \ldots\ldots$

DUAL PROBLEM $= \ldots\ldots$

maximize: $\sum_{i=1}^{N} \alpha_i \underbrace{x_i^T x_i}_{k(x_i, x_i)} - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \underbrace{x_i^T x_j}_{k(x_i, x_j)}$

subject to: $\sum_{i=1}^{N} \alpha_i = 1.$

$0 \leq \alpha_i \leq C \;\; \forall i$

$\underline{\underline{\alpha^*}} \rightarrow \underline{\underline{a^*}} \quad \underline{\underline{R^*}}$

$\Downarrow$

$\|x - a^*\|^2 \leq R^{*2}$

NOT AN OUTLIER $\Leftarrow$ TRUE x is in the circle

AN OUTLIER $\Leftarrow$ FALSE x is outside of the circle