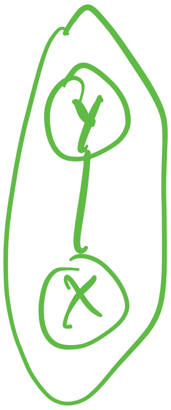


$y \in \{+1, -1\}$

$x \in \mathbb{R}^D$

$\arg\max_{y \in \{\pm 1\}} P(y|x) \propto w^T \Phi(x,y)$

$\Phi(x,y) \Rightarrow [\![y \in r]\!] \wedge \underline{[\![x_5 = \infty]\!]}$

$$\Rightarrow \Phi(x,y) = \begin{pmatrix} \left.\vec{x}\right\} & y == +1 \\ \left.\vec{x}\right\} & y == -1 \end{pmatrix}$$

$\Rightarrow$ logistic regression

$\Rightarrow$ Crammer & Singer, 2007

pseudo examples:

training set: $\boxed{\{(x_i, \cancel{y}\, y_i)\} \in \mathcal{D}}$

y_i = ~~Det~~ Det N V Det N ←

x_i = The dog chases the cat

$\arg\max_y f(x, y)$ → $f(x_i, y_i)$ should be largest

$f(x_i$, NNNNN) is worse

$f(x_i$, NVVDetN) — " —

$$\boxed{\begin{aligned} f(x_i, y_i) &> f(x_i, \text{'NNN...N'}) \\ &\;\;\vdots \\ &> f(x_i, \text{'VNDetDetV'}) \end{aligned}}$$

learning task: find w such that ↙ holds!

$$\boxed{f(x_i, y_i) - f(x_i, NNNNN) > 0}$$

$$f(x_i, y_i) - f(x_i, NNNN\ldots) > 0$$

$$\langle \widetilde{w, \Phi(x_i, y_i)} \rangle + b - \langle \widetilde{w, \Phi(x_i, \bar{y})} \rangle + b > 0 \quad \forall \bar{y} \neq y_i$$

$$\boxed{w^T \left(\Phi(x_i, y_i) - \Phi(x_i, \bar{y}) \right)} > 0 \quad \forall \bar{y} \neq y_i$$

$$\Phi(x_i, y_i) = (0, 1, 1, 0)$$

$$\Phi(x_i, \bar{y}) = (1, 0, 1, 0)$$

$w \leftarrow w + \Phi(\ldots)$

$- \Phi(\ldots)$

$$\Rightarrow \Phi(x_i, y_i) - \Phi(x_i, \bar{y}) = (-1, 1, 0, 0)$$

$$w \leftarrow w + \left(\Phi(x_i, y_i) - \Phi(x_i, \bar{y}) \right)$$

$\forall \bar{y} \neq y_i$

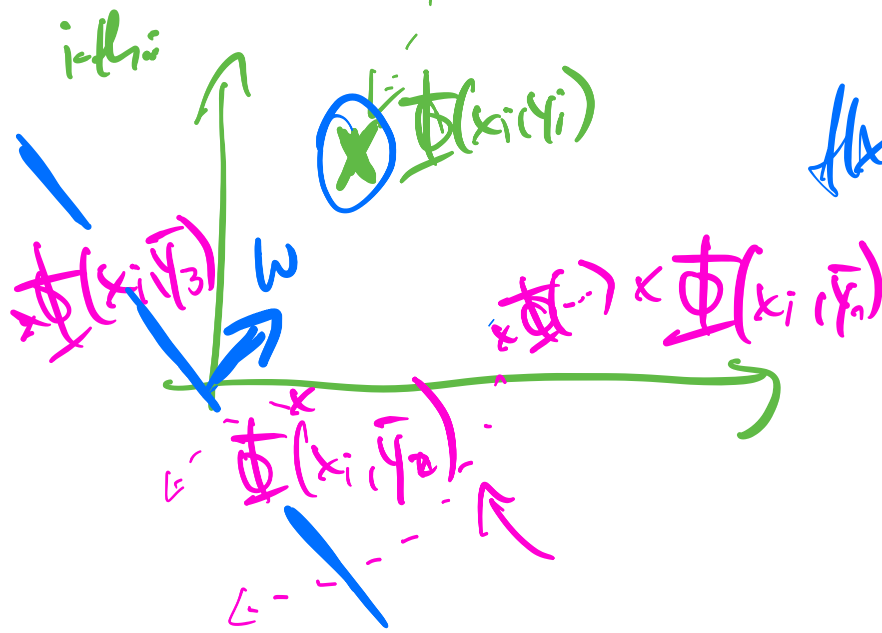
$\Rightarrow f(x, \bar{y}) \stackrel{?}{=} y_i$

$\hat{y} = \arg\max_y$

yes $\rightarrow$ ok

no: update

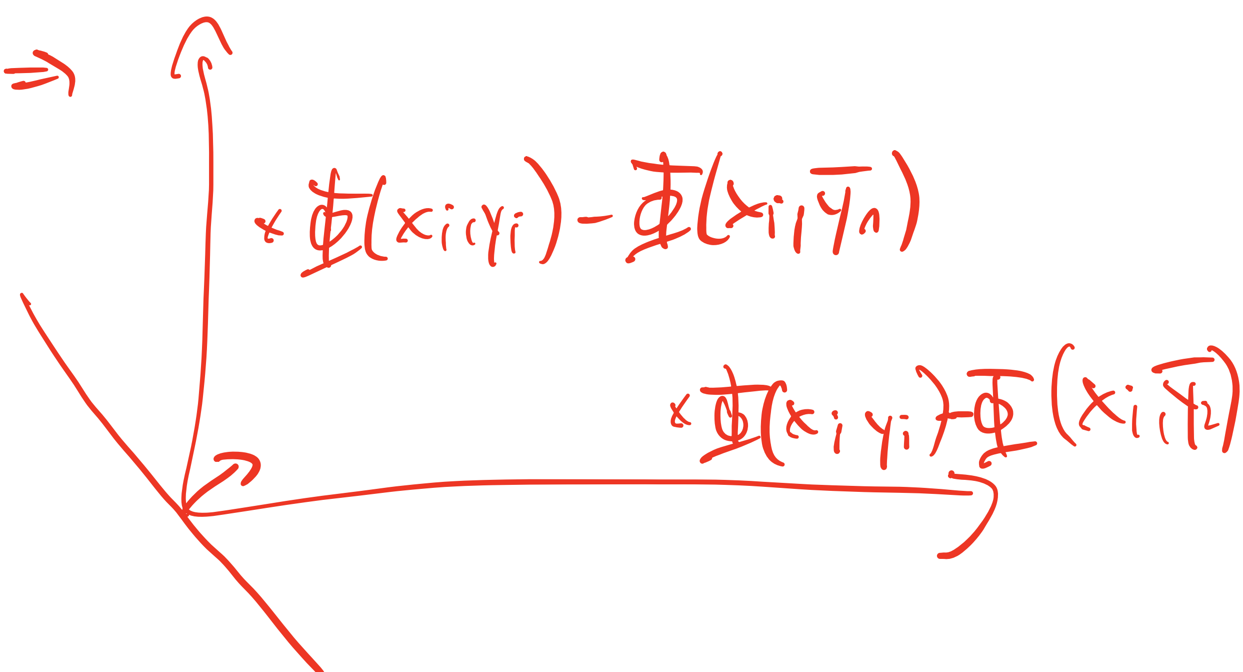
$w \leftarrow \Phi(x_i, y_i) - \Phi(x_i, \hat{y})$



$f(x, y) = w\Phi(x, y)$

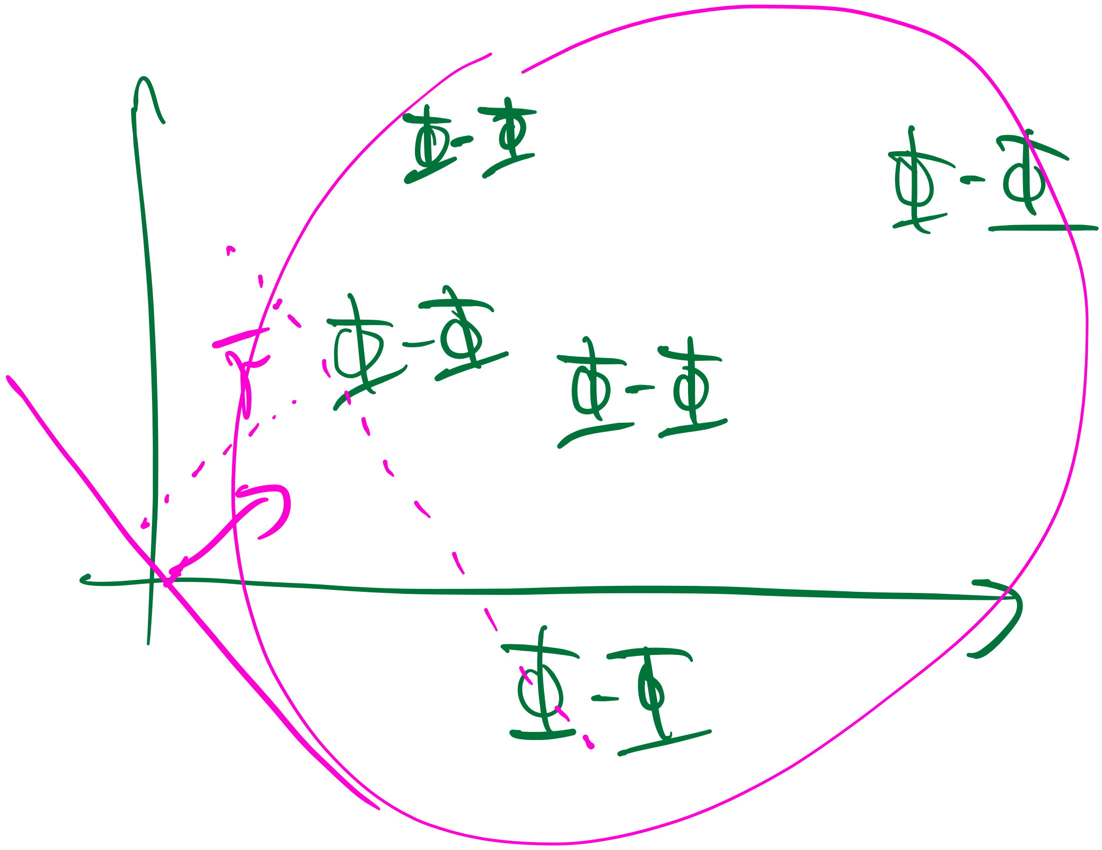
$\Rightarrow$ in dir of $\Phi(x_i, y_i) - \Phi(x_i, \bar{y})$

$\Rightarrow$



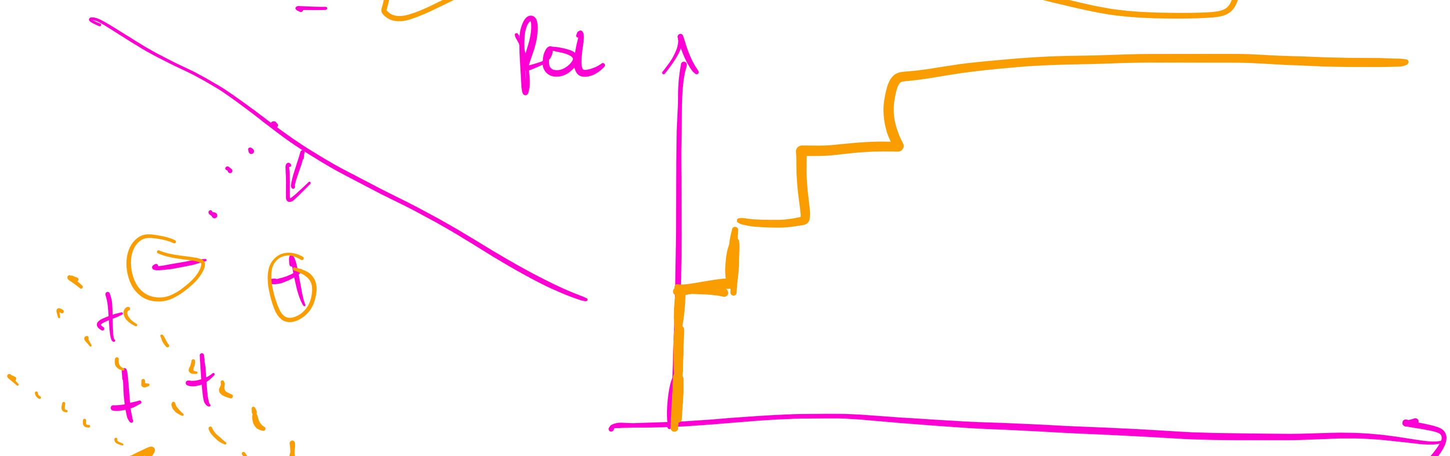
$\Phi(x_i, \cdot)$ is fixed

$\Rightarrow$



connection to AUC

$$\boxed{Pr\left(f(x^+) > f(x^-)\right)}$$



$f(x^+) > f(x^-)$!

$w^T \phi(x^+) + b > w^T \phi(x^-) + b$

$w^T\left(\phi(x^+) - \phi(x^-)\right) > 0$

Training set of size N

$\frac{N}{2}$ positive and

$\frac{N}{2}$ negative ones

Because:

$\left(\frac{N}{2}\right)^2$ quadratic

→ AUC SVM, 2005

A

k

j

a_j