Optimization objective. $\boxed{\text{Slides } 28 \sim 30}$ $y_i(\theta^T x_i + b) - 1 \geq 0.$

$$\tilde{J} = \frac{1}{2}\sum_{j=1}^{d}\theta_j^2 - \sum_{i=1}^{n} a_i\left(y_i(\theta^T x_i + b) - 1\right)$$

$(\frac{1}{2}x^2)' = 2 \cdot \frac{1}{2} \cdot x = x$

$$= \frac{1}{2}\sum_{j=1}^{d}\theta_j^2 - \sum_{i=1}^{n} a_i\left(y_i(\theta^T x_i + b)\right) + \sum_{i=1}^{n} a_i$$

$$= \underline{\frac{1}{2}\sum_{j=1}^{d}\theta_j^2} - \sum_{i=1}^{n}\underbrace{a_i(y_i\theta^T x_i)}_{\text{constants}} - b \cdot \underbrace{\sum_{i=1}^{n} a_i y_i}_{\text{constant}} + \sum_{i=1}^{n} a_i$$

$\to 0$

$\Rightarrow$ want to find the critical pt. $\theta ; b$ = separate / unlike linear / logistic regression.

$\theta = w$

$$\frac{\partial J}{\partial \theta} = \frac{1}{2} \cdot 2 \cdot \theta - \sum_{i=1}^{n} a_i y_i x_i = 0 \text{ (by being the critical pt)}$$

$$\Rightarrow \underline{\theta = \sum_{i=1}^{n} a_i y_i x_i.}$$

$$\frac{\partial J}{\partial b} = 0 - 0 - \sum_{i=1}^{n} a_i y_i + 0 = 0 \text{ (critical pt).}$$

$$-\left(\boxed{-\sum_{i=1}^{n} a_i y_i}\right) = 0 \Rightarrow \underline{\sum_{i=1}^{n} a_i y_i = 0.}$$

$\boxed{\text{Slide 31}}$

$$J = \frac{1}{2}\sum_{i=1}\sum_{j=1} a_i y_i x_i \cdot a_j y_j x_j - \sum_{i=1}\sum_{j=1} a_j y_j x_j \cdot a_i y_i x_i - 0 + \sum_{i=1} a_i$$

measuring the similarity between vecs.

$$= \underline{-\frac{1}{2}\sum_{i=1}\sum_{j=1} a_i a_j y_i y_j x_i x_j + \sum_{i=1} a_i}$$

$h(x) = y = \theta^T x + b$

$\underline{b = y - \theta^T x}.$

incoming $= (x, y)$, know $(x_i, y_i) \cdots (x_{SV}, y_{SV})$

$h(x) = y = \theta^T x + b. \Rightarrow h(x) = \text{sign}(\theta^T x + b).$

the one we want to classify

incoming.

$$= \text{sign}\left(\sum_{i=1}^{SV} a_i y_i x_i \cdot \boxed{x} + b\right) \qquad b = \frac{1}{|SV|}\sum_{SV}\left(y_i - \sum_{i=1}^{SV} a_i y_i x_i \cdot x\right)$$

Support vector

SV's label.

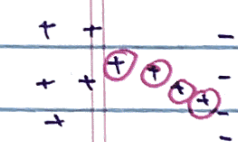
$\boxed{\text{Slide 32}}$

$\sum_{i=1}^{SV} a_i y_i = 0.$ if lots of $y_i = -1$, different

then, the lagrangian multipliers (a_j) for $y_j = +1$ must be large.

+

$\boxed{\text{Slide 35}}$



$\min_{\theta} \frac{1}{2} \sum_{j=1}^{d} \theta_j^2 + \underset{\text{Coeff}}{\underline{\underline{C}}} \cdot \sum_i \varepsilon_i \rightarrow$ each ε_i corresponds to $\underline{\underline{x_i}}$ a vector in training.

$\boxed{\text{Slide 42}}$

Similarity $\langle x_i, x_j \rangle$. $\rightarrow$ cosine similarity

$\Phi[x_{i1}, x_{i2}] = [x_{i1}, x_{i2}, x_{i1} \cdot x_{i2}, x_{i1}^2, x_{i2}^2]$.

$\Phi[x_{i1}, x_{j2}] = [x_{j1}, x_{j2}, x_{j1} \cdot x_{j2}, x_{j1}^2, x_{j2}^2]$

$x_i = \underbrace{[x_{i1}, x_{i2}]}_{\text{2-dimensionality vector}}$.

instead of $C(x_i, x_j)$; you compare $(\Phi(x_i), \Phi(x_j))$

$\Phi = x \rightarrow \hat{x} = \Phi(x)$ $(a+b)^2 = a^2 + 2ab + b^2$ there's no Φ

$k(x_i, x_j) = (x_{i1} \cdot x_{j1} + x_{i2} \cdot x_{j2})^2$

$= (x_{i1}^2 x_{j1}^2 + 2 x_{i1} x_{i2} x_{j1} x_{j2} + x_{i2}^2 \cdot x_{j2}^2)$.

claim $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ equivalent

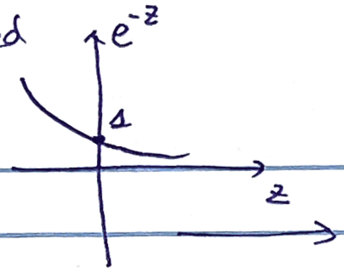
$\Phi(x_i) = [x_{i1}^2, x_{i2}^2, \sqrt{2}\, x_{i1} \cdot x_{i2}]$

$\Phi(x_j) = [x_{j1}^2, x_{j2}^2, \sqrt{2}\, x_{j1} \cdot x_{j2}]$

$\boxed{\text{Slide 43}}$

not squared

$k(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right)} = e^{-z}$; $z = \frac{\|x_i - x_j\|_2^2}{2\sigma^2}$



= if $x_i = x_j$; $z = \frac{0}{2\sigma^2} = 0$; $k(x_i, x_j) = 1$.

$k(x_i, x_j) \in [0, 1]$

if $x_i \neq x_j$ $\quad z = \frac{\infty}{2\sigma^2}$; $k(x_i, x_j) = e^{-z} = 0$

$x_i - x_j = -\infty$

vector x. $\|x\|$ = magnitude of x = $\sqrt{x_{i1}^2 + x_{i2}^2}$

$= \underline{\|x\|_2^2 = \|x\|^2 = \|x\|_2}$

$\underline{(\|x\|)^2}$ = squared magnitude $\quad \|x\|_2^2$