Into
What svm DO?
Linear svm

Linear svm
→ Cat / Non cat
Train → $(x_1 y_1), \ldots (x_n y_n)$
Hyperplane
Hypeplen → $w^T x - b = 0$

Hard margin
(face Recog)
Hyperplne
$w^T x_i - b \geq 1$ anythg on above the boundar is 1

Distance betwn two hyperpln $\frac{2}{\|w\|}$
min $\|w\|$ $(y_i (w^T x_i - b) \geq 1)$

$x^5 + 2x^3y^2 + 9xy^4$

Soft max

$$\frac{1}{n} + \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i - b)) + \lambda \|w\|^2$$

Non-linear

Poly (homo) : $K(\vec{x_i}, \vec{x_j}) = (\vec{x_i} \cdot \vec{x_j})^d$

(inhom) : $\quad \text{''} \quad \text{''} = (\vec{x_i} \cdot \vec{x_j} + 1)^d$

RBF $= K(\vec{x_i}, \vec{x_j}) = \exp(-\lambda \|\vec{x_i} - \vec{x_j}\|^2)$

$\lambda > 0 \quad \& \quad \lambda = \frac{1}{2 \times (\sigma)^2}$

$\vec{w} \cdot \phi(x) = \sum_{i=1} \alpha_i y_i k(x_i, x_j)$

Math

$$\frac{1}{n} + \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i - b)) + \lambda \|w\|^2$$

Primal Problem

$w^T x_i - b$

for each $i \in (1, \ldots, n)$, $\varepsilon_i = \max(0, 1 - y_i(\ ))$

$y_i(w^T x_i - b) \geq 1 - \varepsilon_i \rightarrow$ smaller non m

$$\min \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i + \lambda \|w\|^2$$

sub $\rightarrow$ $y_i(w^T x_i - b) \geq 1 - \varepsilon_i$ x

$\varepsilon_i \geq 0$, for all i

Dual Problem

$$\max f(c_1 \ldots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (x_i^T x_j) y_j c_j$$

Subject to $\sum_{i=1}^{n} c_i y_i = 0$ & $0 \leq c_i < \frac{1}{2n\lambda}$ for all i

$b = w^T x_i - y$

Sub gradient

$$f(w, b) = \left[\frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i - b)) + \lambda \|w\|^2\right]$$

Coordinate descent

$\max f(c_1, \dots c_n)$

$$= \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (x_i \cdot x_j) y_j c_j$$

Empirical Risk mmt

$X_1, \dots x_n \quad y_1 \dots y_n$

y_{n+1} given x_{n+1}

$$\mathcal{E}(f) = E[l(y_{n+1}, f(x_{n+1}))$$

$$\mathcal{E}(f) = \frac{1}{n} \sum_{i=1}^{m} l(y_i f(x_n))$$

~~Transductive SVM~~

Regressi

$\min \frac{1}{2} \|\omega\|^2$

Subject $|y_i - \langle \omega, x_i \rangle| \le \epsilon$

Page No:
Date:
YOUVA

Kernel Trick

- → Write an algorithm in terms of $\langle x, z \rangle$
- → We are mapping $x \to \varphi(x)$
- → Find a way to compute $K(x,z) = \varphi(x)^T \cdot \varphi(z)$
- → Replace x, z from algo. with $K(x,z)$

* Example :-

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \longrightarrow \varphi(x) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ x_1 x_3 \\ x_2 x_1 \\ x_2 x_2 \\ x_2 x_3 \\ x_3 x_1 \\ x_3 x_2 \\ x_3 x_3 \end{bmatrix}$$

$$x \in \mathbb{R}^n \longrightarrow \phi(x) \in \mathbb{R}^{n^2}$$

$$\phi(z) = \begin{bmatrix} z_1 z_1 \\ z_1 z_2 \\ z_1 z_3 \\ z_2 z_1 \\ z_2 z_2 \\ z_3 z_3 \\ z_3 z_1 \\ z_3 z_2 \\ z_3 z_3 \end{bmatrix}$$

* We are taking $O(n^2)$ time to compute $\phi(x)$ or $\phi(z)$

$$k(x,z) = \phi(x)^T \phi(z) = (x^T z)^2$$

$$= \left(\sum_{i=1}^{n} x_i z_i\right)\left(\sum_{j=1}^{n} x_j z_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i z_i x_j z_j \quad \boxed{O(n)}$$

Proved! $\phi(x)^T \phi(z)$

How to make Kernels

If x, z are "similar" the $k(x,z) = \phi(x)^T \phi(z)$ is large

If x, z are dissimilar then $k(x,z)$ is "small"

$$K(x,z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right)$$