EECS 127

Machine Learning Applications.

Admin: Find project partners! Post to Piazza + Slack.

Classification problem. : Support Vector Machines (SVMs).

 cat or dog.

Bail or not, Disease or not ... loan or not ..

$(\vec{x}, y)$ $\quad y \in \{+1, -1\}$

feature vector, label

Training data: $(\vec{x}_1, y_1), (\vec{x}_2, y_2) \cdots (\vec{x}_n, y_n)$

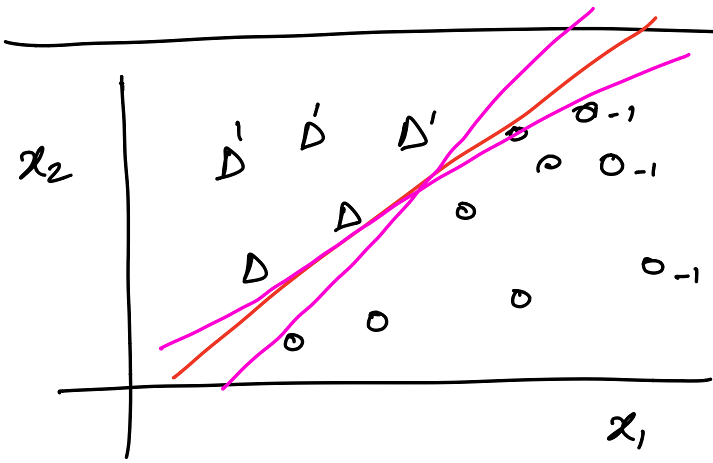
n data points

eg. cat pictures v/s dog pictures.

$\vec{x}_{new}$: want to find label

Logistic Regression.

Linear Regression:
$$\begin{bmatrix} - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_n & - \end{bmatrix} \begin{bmatrix} \uparrow \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
$\vec{w}$



Find: A separating hyperplane that separates the two classes.

Want to find: $f(\vec{x}) = \vec{w}^T \vec{x} - b$ such that

$$\boxed{\begin{aligned} f(\vec{x}_i) &= \vec{w}\,\vec{x}_i - b > 0 \quad \text{if } y_i = +1 \\ f(\vec{x}_i) &= \quad\quad\quad\quad < 0 \quad \text{if } y_i = -1 \end{aligned}}$$

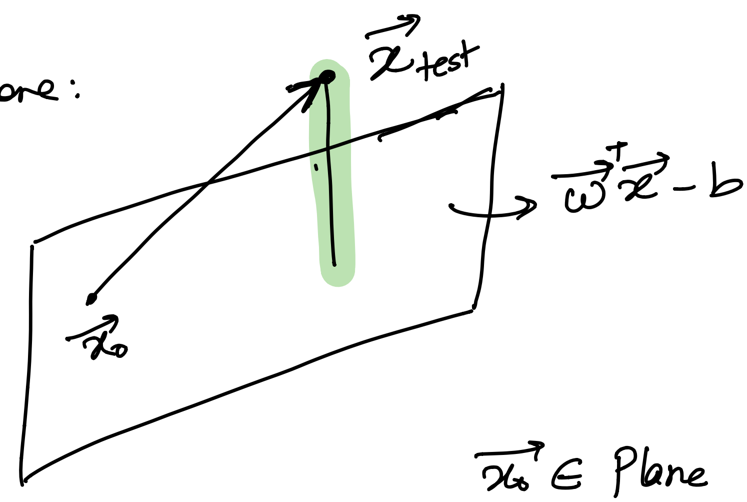
"Best" linear separator:

How far away are points from the separator?

"MARGIN" : Want to find a separator with maximum margin.

→ Distance of the closest point to the hyperplane.

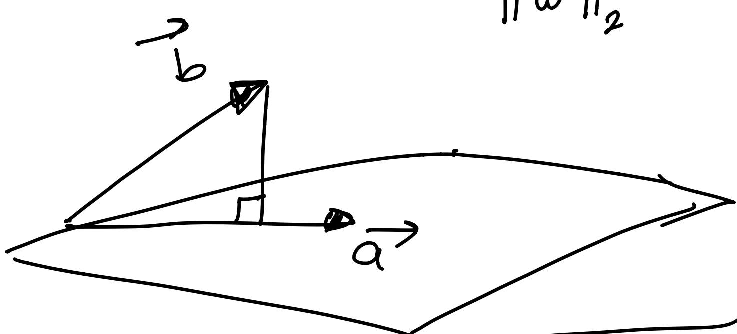
Hyperplane:



$\vec{x}_0 \in$ Plane

$\vec{x}_0 \in$ Plane

Projection: $\dfrac{\vec{w}^T(\vec{x}_{test} - \vec{x}_0)}{\|\vec{w}\|_2}$



$\vec{a} \cdot x = \vec{b}$

Find the best $\vec{x}$?

$(A^TA)^{-1}A^T\vec{b}$

$(\vec{a}^T\vec{a})^{-1}\vec{a}^T\vec{b} = \dfrac{\vec{a}^T\vec{b}}{\|\vec{a}\|_2^2}$

Distance: $\dfrac{\vec{w}^T(\vec{x}-\vec{x_0})}{\|\vec{w}\|_2}$ } margin

---

$$m = \frac{\vec{w}^T(x-\vec{x_0})}{\|\vec{w}\|_2}$$ no square.

$$\max_{\vec{w},\, b,\, m} \; m.$$

s.t. $y_i(\vec{w}^T\vec{x_i} - b) \geq 0 \quad \forall i$ Classify Correctly.

$$\underbrace{\frac{|\vec{w}^T\vec{x_i} - \vec{b}|}{\|\vec{w}\|_2}}_{\text{distance}} \geq \underbrace{m}_{\text{margin}} \quad \forall i$$

→ If $y_i = +1$, $\vec{w}^T\vec{x_i} - b > 0$

If $y_i = -1$, $\vec{w}^T\vec{x_i} - b < 0$

$$|\vec{w}^T\vec{x_i} - b| = y_i(\vec{w}^T\vec{x_i} - b)$$

$$\boxed{\begin{aligned} &\max_{\vec{w}, m, b} \quad m \\ &\frac{y_i(\vec{w}_i^T \vec{x}_i - b)}{\|\vec{w}\|_2} \geq m \quad \forall i \\ &m \geq 0 \end{aligned}}$$

Hard-margin SVM.

$$\vec{w}^T x - b = 0$$

$$\alpha \vec{w}^T \vec{x} - \alpha \vec{b} = 0.$$

If $(m, \vec{w}, b)$ is a solution to the optimization, then $(m, \alpha\vec{w}, \alpha b)$ is also a solution.

$$y_i \frac{(\alpha\vec{w}^T \vec{x}_i - \alpha b)}{\alpha \|\vec{w}\|_2} \geq m$$

One way to solve this: $\|\vec{w}\|_2 = 1$

Choose: $\|\vec{w}\|_2 = \frac{1}{m}$ ←

Rewrite constraint:

$$y_i(\vec{w}^T \vec{x}_i - b) \geq 1$$

Program: $\max \left(\|\vec{w}\|_2\right)^{-1}$

$$\text{s.t. } y_i(\vec{w}^T \vec{x} - b) \geq 1 \quad \forall i$$

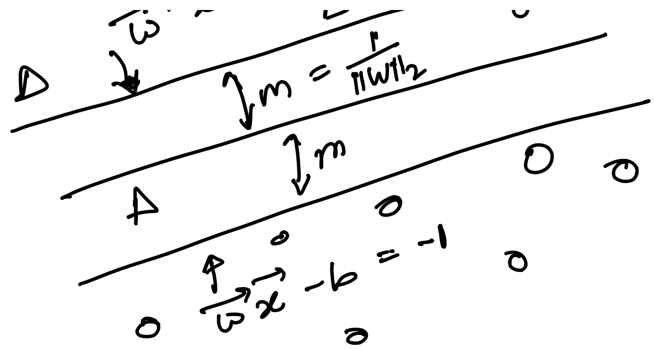
$$\boxed{\begin{aligned} &\min \tfrac{1}{2}\|\vec{w}\|_2^2 \quad \leftarrow \text{Quadratic} \\ &\text{s.t. } y_i(\vec{w}^T \vec{x}_i - b) \geq 1 \quad \forall i \leftarrow \text{Linear} \end{aligned}}$$

Standard form of the ~~two~~ hard margin SVM.

QP.

$\vec{w}^T \vec{x} - b = 1$

$m = \frac{1}{\|\vec{W}\|_2}$



$\frac{2}{\|\vec{\omega}\|_2}$ distance between the closest points on either side.

Soft margin SVM.

Relax: $y_i(\vec{\omega}^T\vec{x}_i - b) \geq 1$

$$\implies \quad y_i\left(\vec{\omega}^T\vec{x}_i - b\right) \geq 1 - \xi_i$$

slack variable

$$\xi_i \geq 0$$

$$\min_{\vec{\omega},\, \vec{b},\, \xi_i} \quad \frac{1}{2}\|\vec{\omega}\|_2^2 + C \cdot \sum_{i=1}^{n} \xi_i$$

Regularizer

$$\text{s.t.} \quad y_i(\vec{\omega}^T\vec{x}_i - b) \geq 1 - \xi_i \quad \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

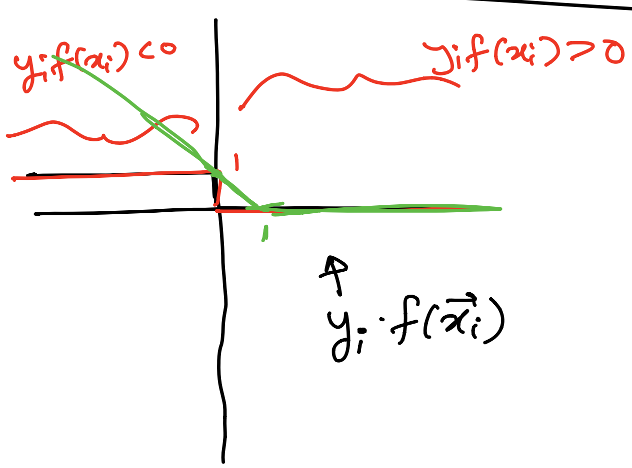
C : Hyperparameter.

Small C : less sensitive to violations of margin

Large C : Super sensitive to violation of margin.

$\boxed{\text{Soft - Margin SVM.}}$

Hinge - loss formulation



$$L_{01}(y_i, f(x_i)) = \begin{cases} 0 & y_i f(x_i) > 0 \\ 1 & y_i f(x_i) < 0 \end{cases}$$

$$\min \cdot \frac{1}{n} \sum_{i=1}^{n} L_{0-1}(y_i, \vec{w}\vec{x}_i - b) \quad \text{Non-convex}$$

$$L_{hinge}(y_i, f(x_i)) = \max\left(0, 1 - y_i \cdot f(\vec{x}_i)\right)$$

$$\min_{\vec{w}, b} \frac{1}{n} \sum_{i=1}^{n} L_{hinge}(y_i, \vec{w}^T \vec{x}_i - b) + \underbrace{\lambda \|\vec{w}\|_2^2}_{\text{regularizer.}}$$

Rewrite soft-margin SVM.

$$\min \frac{1}{2} \|\vec{w}\|_2^2 + C \cdot \sum_{i=1}^{n} \xi_i$$

$$\xi_i \geq \max(0, 1 - y_i(\vec{w}^T \vec{x}_i - b))$$

$$\min \frac{1}{2} \|\vec{w}\|_2^2 + C \sum_{i=1}^{n} \xi_i$$

$$\xi_i = \max(0, 1 - y_i(\vec{w}^T \vec{x}_i - b))$$

$$\min_{\vec{w}, \vec{b}} \frac{1}{2} \|\vec{w}\|_2^2 + C \cdot \sum_{i=1}^{n} \max(0, 1 - y_i \cdots)$$

Regularized hinge loss problem.