Classification

$y \in \mathbb{R}$ $-\infty, \infty$ $0, \infty$

x y If y is categorical

y is a class label

We will assume that y takes only 2 values.

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

Credit card approval system:

y → accept $(+1)$
y → reject (-1)

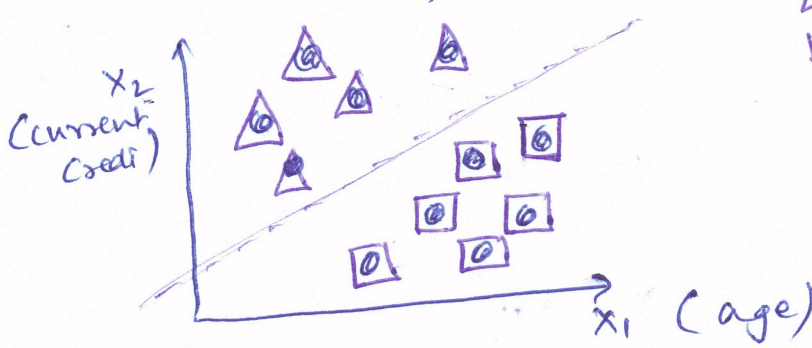
x (age, current credit)
x_1, x_2

Some function f : $f(x) = y$

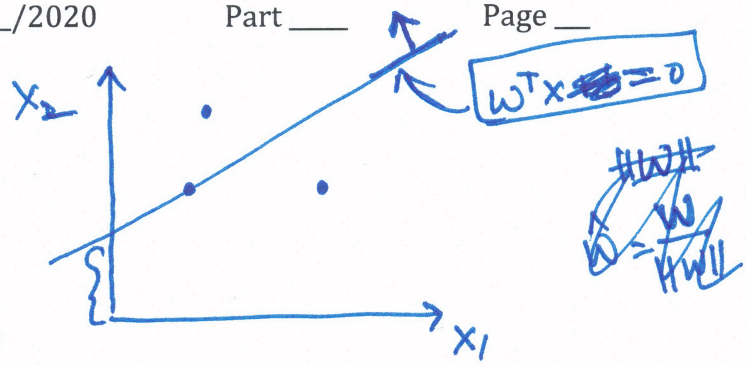
decision function

Inductive bias : f is a line

△ - reject
□ - accept.



↳ linear classifiers



$$x_2 = m x_1 + c$$

$$\boxed{-c - m x_1 + x_2 = 0}$$

$$(-c)\cdot 1 + (-m) x_1 + (1) x_2 = 0$$

are the same

$$\boxed{\begin{bmatrix} -c \\ -m \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0}$$

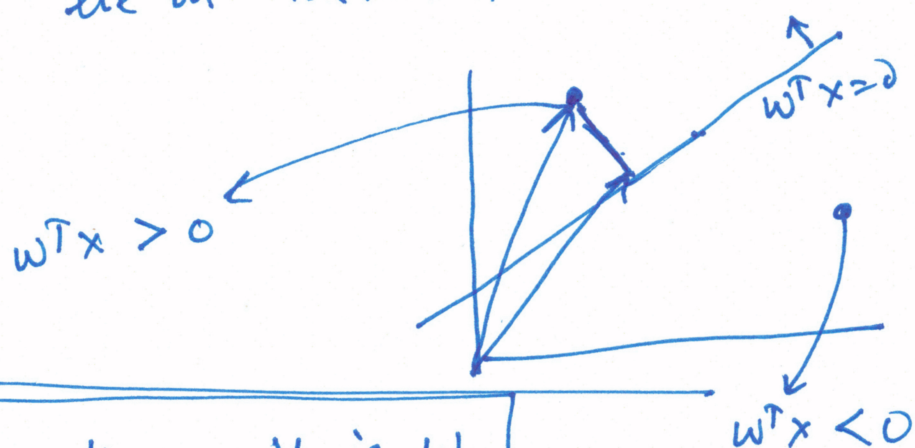
~~$w^T x + b = 0$~~

we have absorbed intercept term into **w** and **x**

$$\boxed{w^T x = 0}$$

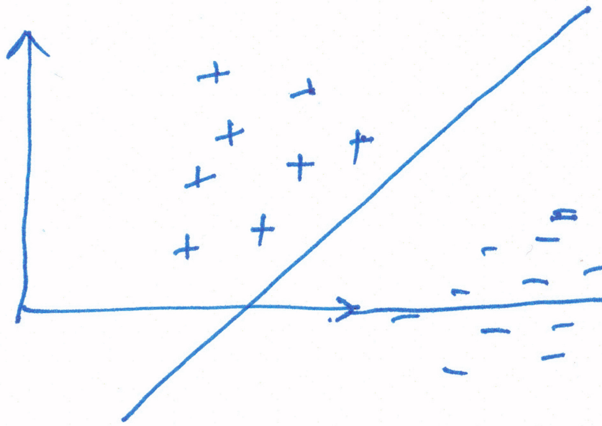
If **x** lies on the line: $w^T x = 0$

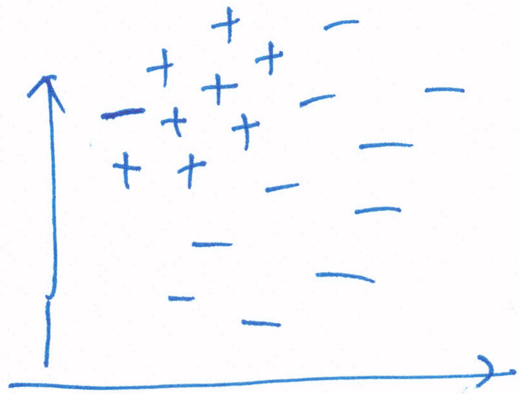
If x does not lie on the line: $w^T x \neq 0$



$$\boxed{\begin{array}{llll} \text{If} & w^T x \geq 0 & \text{then} & y \text{ is } +1 \\ \text{If} & w^T x < 0 & \text{then} & y \text{ is } -1 \end{array}}$$

How do we learn w?





NBT

Define an objective fn. $J(w)$ using X, y

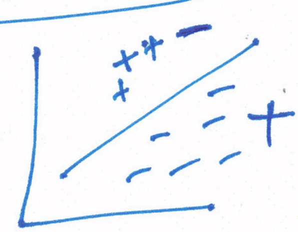
Depending on J → Perceptron
→ Logistic Regression.
→ Support Vector Machine

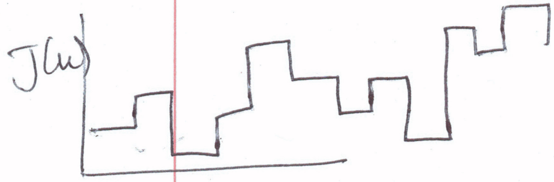
Ideal objective fn.

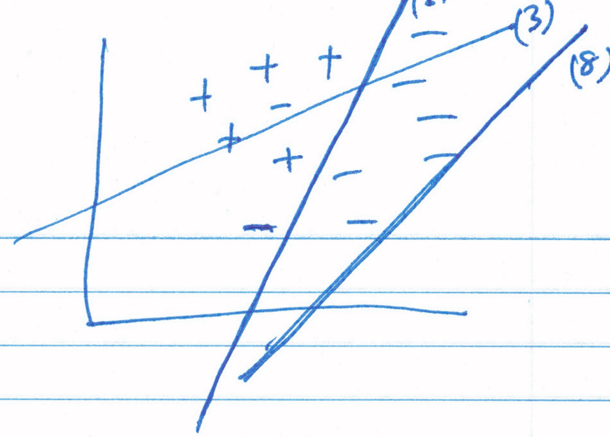
$$\boxed{J(w) = \sum_{i=1}^{N} \mathbb{I}[y_i(w^T x_i) < 0]}$$

counting number of mistakes.

0-1 loss



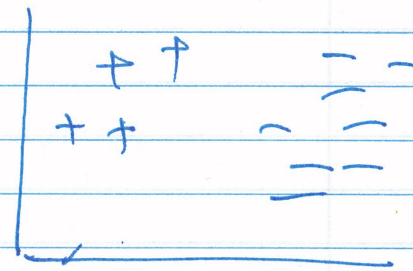




Perceptron

$$\boxed{J(w) = \frac{1}{2}\sum_{i=1}^{N} (y_i - w^T x_i)^2}$$

squared loss.



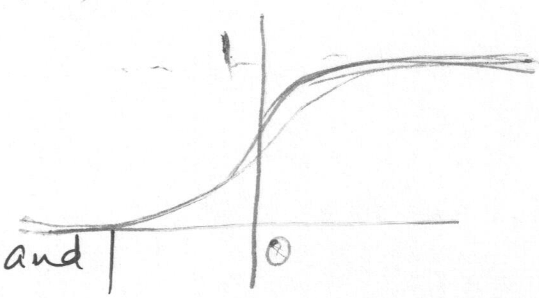
Logistic Regression

$\mathbf{w} \longrightarrow$ D+1 length vector.

$\mathbf{w}^T\mathbf{x} \longrightarrow$ D+1 length vector.

$$\sigma(w^T x) = \frac{1}{1+\exp(-w^T x)}$$

sigmoid fn.



$\sigma(w^T x)$ will be between 0 and 1

If $w^T x > 0$ then $y = +1$

This is same as saying

$$\frac{1}{1+\exp(-w^T x)} \geq 0.5 \quad \text{then } y = +1$$
$$\text{else} \quad y = -1$$

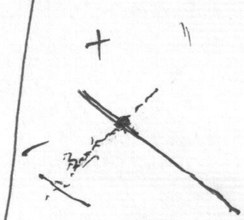
If $w^T x = 0 \Rightarrow \sigma(w^T x) = 0.5$

If $w^T x > 0 \Rightarrow \sigma(w^T x) > 0.5$

If $w^T x < 0 \Rightarrow \sigma(w^T x) < 0.5$

$$\sigma(w^T x) = \frac{1}{1+\exp(-w^T x)} \equiv P(y=+1)$$

$$1-\sigma(w^T x) \qquad P(y=-1)$$



$$J(w) = \sum_{i=1}^{N} \log(1 + \exp(-y_i w^T x_i))$$

$$\boxed{\frac{d}{dw} J(w) = \sum_{i=1}^{N} \frac{d}{dw} \left(\qquad \right)}$$

Hessian

$J(w)$

$$\nabla \quad \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \vdots \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial}{\partial w_1}\frac{\partial J}{\partial w_1} & \frac{\partial}{\partial w_1}\frac{\partial J}{\partial w_2} & \cdots \\ & & \end{bmatrix}$$