Machine learning

A What is ML?

→ A set of methods for making decisions from data.

• why study ML?

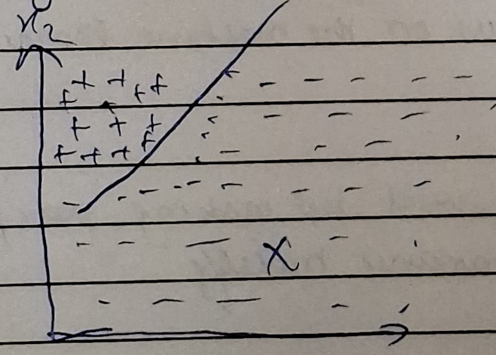
→ To apply; to understand; to evaluate.

* We have training data.

For data point $i \in \{1, \dots n\}$

- Feature vector
$$x^{(i)} = (x_1^{(i)}, \dots x_d^{(i)})^T \in R^d$$

- label $y^{(i)} \in \{-1, +1\}$



Training Data $D_n = \{(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})\}$

* what we want? A good way to label new pts.

7044929440

classmate

Date

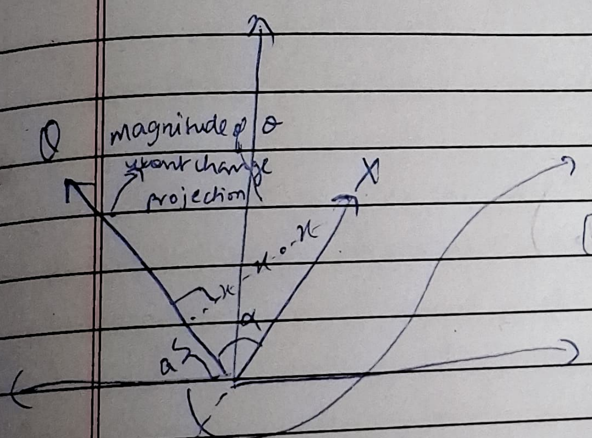
Page

* Let $h: R^d \to \{-1, +1\}$

$x \to h \to y$

* Linear Classifiers

Example H: All hypotheses that label +1 on one side of a line -1 on the other side.



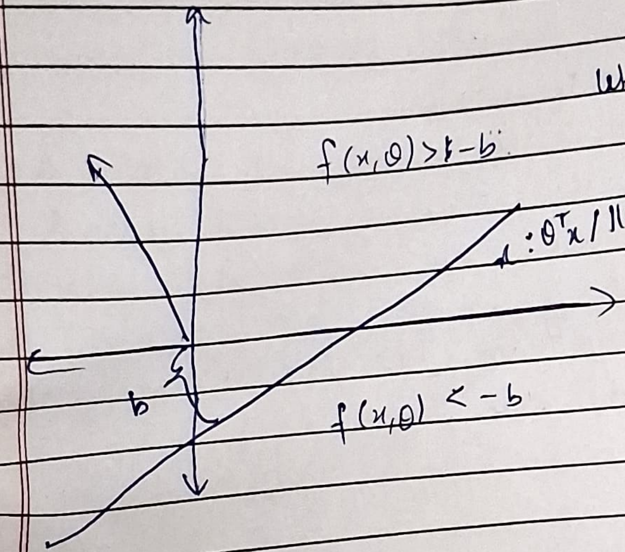
$\theta^T x / \|\theta\|$ (Projection of x onto θ)

$[1 \times d] \times [d \times 1]$ dimensions

if $\alpha = 90°$, $|x|$ doesn't matter
projection will be 0

$x: \theta^T x / \|\theta\| = a$ will be the dotted line.

let $x: \theta^T x / \|\theta\|$ be $f(x, \theta)$



$= x: \theta^T x + b\|\theta\| = 0$

$x: \theta^T x + \theta_0 = 0$

$\theta_0 = b\|\theta\|$

- linear classifier

$h(x) = \text{sign}(\theta^T x + \theta_0)$

$$= \begin{cases} +1 & \text{if } \theta^T x + \theta_0 > 0 \\ -1 & \text{if } \theta^T x + \theta_0 \leq 0 \end{cases}$$

$h(x; \theta, \theta_0)$

↑↗

parameters (not inputs)

H = set of all h

The double bars on $||\theta||$ represent a norm (vector)
It calculates the magnitude of that vector in multi dimensional space

How good is a classifier

- It should predict well on future data.
- How good is a classifier at a single point? Loss $L(g,a)$
  g: guess ; a: actual.

- Our guess should be closer to the actual value.

Example: 0-1 Loss.

$$L(g,a) = \begin{cases} 0 & \text{if } g=a \\ 1 & \text{else.} \end{cases}$$

: Asymmetric loss.

$$L(g,a) = \begin{cases} 1 \text{ if } g=1, \; a=-1 & \text{(Newborn didn't have seizure, diagnosed else)} \\ 100 \text{ if } g=-1, \; a=1 & \text{(Newborn had seizure, diagnosed else)} \\ 0 \text{ else.} & \text{(correct diagnosis)} \end{cases}$$

- Test error (n' new points) $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} \left(L(h(x^{(i)}), y^{(i)}\right)$

- Training error $\mathcal{E}_n(h) = \frac{1}{n} \sum_{n=1}^{n} L(h(x^{(i)}), y^{(i)})$

- Prefer h to $\tilde{h}$ if $\mathcal{E}_n(h) < \mathcal{E}_n(\tilde{h})$

Learning a classifier.

Recall: $x \to \boxed{h} \to y$

New:

$$D_n \longrightarrow \boxed{\text{Learning Algorithm}} \longrightarrow h$$

Ex

for $j = 1, \dots$ 1 million
  Randomly sample $(\theta^{(j)}, \theta_0^{(j)})$
  Set $h^{(j)}(x) = h(x; \theta^{(j)}, \theta_0^{(j)})$

Ex — learning-alg (D_n; k < 1 trillion)
hyperparameter
  Set $j^* = \arg\min_{j \in \{1, \dots, k\}} \mathcal{E}_n(h^{(j)})$
  Return $h^{(j^*)}$