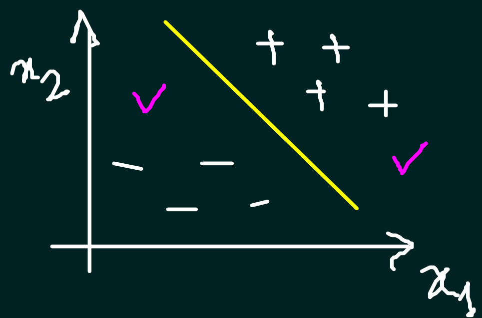
▷ Linear Classification

$[x_0 = 1]$

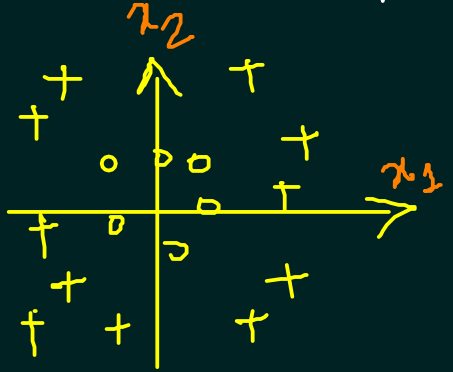
$$\sum_{i=0}^{d} w_i x_i = W^T X \longrightarrow$$



$$\text{sign}(W^T X) \begin{cases} \nearrow +1 \\ \searrow -1 \end{cases}$$

$\uparrow$ new

linearly separable point



NOT linearly separable

↳ How we apply linear classification?

▷ Linear Regression

$$\sum_{i=0}^{d} w_i x_i = W^T X$$

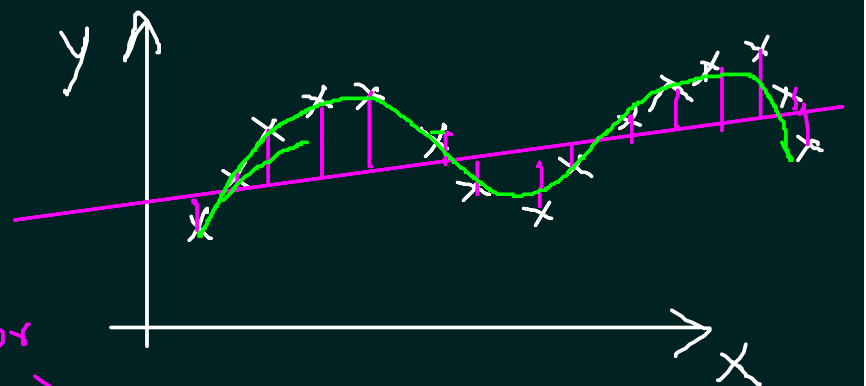
each(x_i, y_i)

$$E_{in}(w) = \frac{1}{N} \sum_{i=1}^{N} (y_i - W^T x_i)^2$$

$$\nabla E_{in}(w) = 0 \Rightarrow w = [(X^T X)^{-1} X^T]\, y$$

$$X = \begin{bmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_N^T & — \end{bmatrix} \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

opt. prob (error Min)



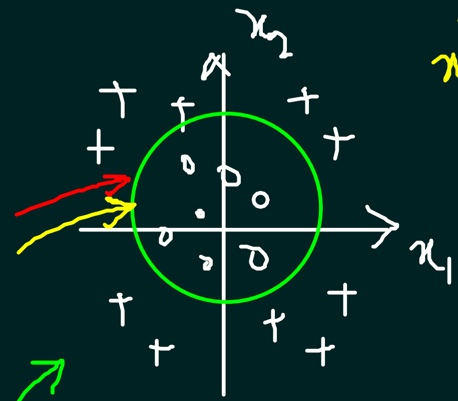
↳ Ramification

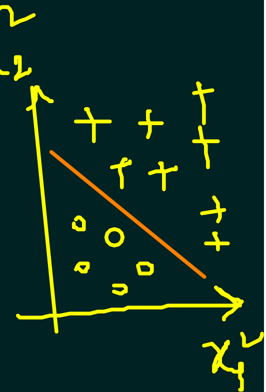
Variant of Linear Classification

Handling Non-linearity

$x \to z = \phi(x)$

$\text{sign}(\tilde{w}^T z) \longrightarrow \tilde{w}^T \text{ to } w$





$(1, x_1, x_2) \xrightarrow{\phi} (1, x_1^2, x_2^2)$

$\to (1, x_1, x_2) \xrightarrow{\phi} (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$ most generic

$\begin{array}{ccc} w_0 & w_1 & w_2 \\ 1, & x_1^2, & x_2^2 \end{array}$?

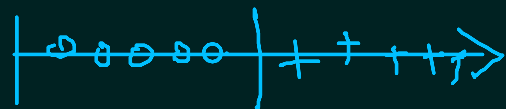
w_0 ✓ $(x_1^2 + x_2^2 - 0.6)$

> 0 ?

< 0

$w_0 \quad w_1$

$(1, x_1^2 + x_2^2)$?

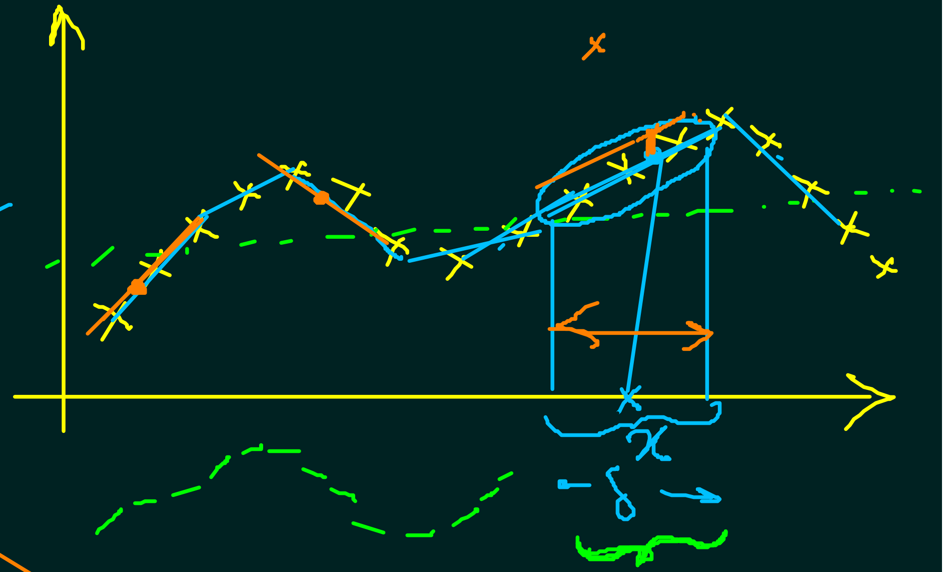


You are learning from data which m/c needs to do

(DATA SNOOPING)

Variant of Linear Regression:

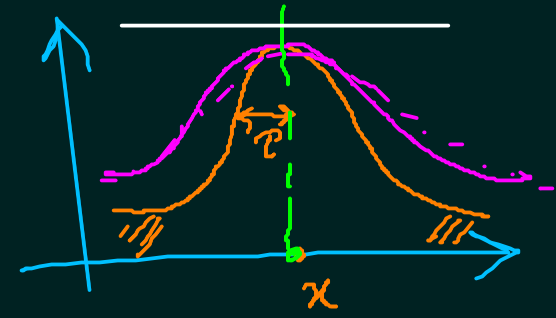
Fit w : Min $\sum_{i=1}^{N} \alpha_i (w^T x_i - y_i)^2$



where, $\alpha_i = \exp\left(-\frac{(x_i - x)^2}{2\tau^2} \right)$

$\forall i$ $|x_i - x| \to$ large $\quad \alpha_i \approx 0$

$|x_i - x| \to$ small $\quad \alpha_i \approx 1$



Locally Weighted Regression

used for noisy data

$f: X \to Y \iff \text{Prob}(y \mid x)$

▷ Predict Heart Attack

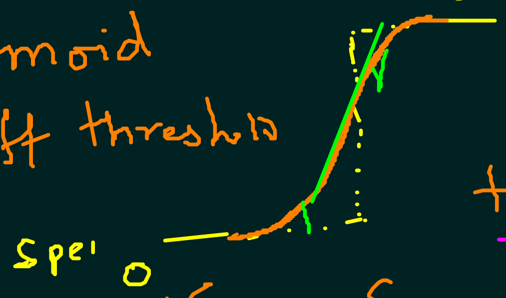
0 — 1

$s = \sum_{i=0}^{d} w_i x_i = w^T x \longrightarrow \theta(s) \longrightarrow h$

$h(x) = \theta(s), \; s = w^T x$

sigmoid

soft threshold



tanh(x)

▷ Logistic Regression:

$\overset{\theta}{(x_1, y_1)} \;\overset{\theta}{-} \; \overset{(1-\theta)}{-}\; - - - \; (x_N, y_N)$ ← TE

$h(x)$

likelihood (MLE)

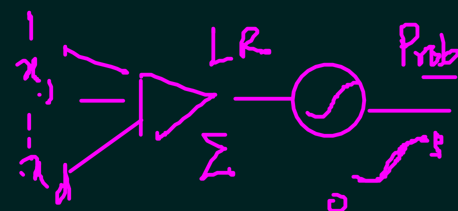
$\theta(s) = \dfrac{e^s}{1+e^s} = \dfrac{1}{1+e^{-s}}$

$$\text{Prob}(y \mid x) = \begin{cases} \overset{\theta(s)}{\cancel{h(x)}}, & y = +1 \\ 1 - \underset{\theta(-s)}{\cancel{h(x)}}, & y = -1 \end{cases}$$

$\theta(-s) = 1 - \theta(s)$

TE, $P(y \mid x) = \theta(y \cdot w^T x)$

$$\prod_{i=1}^{N} P(y \mid x_i) = \prod_{i=1}^{N} \theta(y_i w^T x_i) \longrightarrow \text{MLE}$$



Maximize $\prod_{i=1}^{N} \theta(y_i w^T x_i)$ $\longrightarrow$ Minimize $-\frac{1}{N} \ln \prod_{i=1}^{N} \theta(y_i w^T x_i)$

$$\left[\theta(s) = \frac{1}{1+e^{-s}}\right]$$

$$\text{Min} \; -\frac{1}{N} \sum \ln\left(\frac{1}{1+e^{-y_i w^T x_i}}\right)$$

Logistic Regr.

$$E_{in}(w) = \text{Min}\left[\frac{1}{N}\sum_i \underbrace{\ln\left(1+e^{-y_i w^T x_i}\right)}_{e(h(x_i),\, y_i)}\right]$$

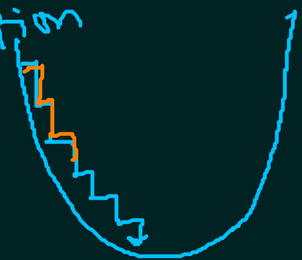
(Iterative)

Linear Regr

$$E_{in}(w) = \text{Min}\left[\frac{1}{N}\sum_i \underbrace{\left(y_i - w^T x_i\right)^2}_{\text{Sum-sq. Error}}\right]$$

$$w = (X^T X)^{-1} X^T y \Leftarrow \text{(One-Step Solution)}$$

Direction



TE$_1$. T.E$_2$.

$\nabla E_{in}(w)$

Itr. :

$$w_i(t) \leftarrow w_i(t-1) - \eta \hat{v}$$

Stepsize

$$\hat{v} = \frac{\nabla E_{in}(w)}{\|E_{in}(w)\|}$$

$$\frac{\partial E_{in}(w)}{\partial w_i} = \left[\frac{\partial E_{in}}{\partial w_0}, \frac{\partial E_{in}}{\partial w_1}, \ldots\right]$$

error

$\eta = \text{small}$ $\qquad$ $n = \text{big}$ $\qquad$ $n = \text{adaptive}$

$$w(t) \leftarrow w(t-1) - \eta \frac{\nabla E_{in}(w)}{\|E_{in}(w)\|}$$

[w initialized to Random]

$$\nabla E_{in}(w) = \overbrace{\quad\quad}$$

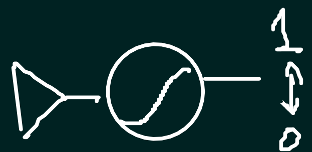
$$\nabla E_{in}(w) = \left[\frac{\partial E_{in}}{\partial w_0}, \frac{\partial E_{in}}{\partial w_1}, \ldots, \frac{\partial E_{in}}{\partial w_d}\right]$$

$$\frac{1}{N} \sum \frac{-y_{i0} x_{i0}}{1 + e^{-y_n w^T x_i}}$$

$$E_{in}(w) = \frac{1}{N} \sum_i \ln\left(\underbrace{1 + e^{-y_i w^T x_i}}\right)$$

Cross entropy error

Lin Class



Logistic Regr.

[Stochastic gradient Descent]

LR $\qquad s = w^T x$