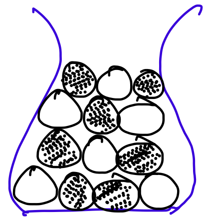
Possible — No
vs
Probable — Yes



$\mu \leftarrow$ Prob(Dark)

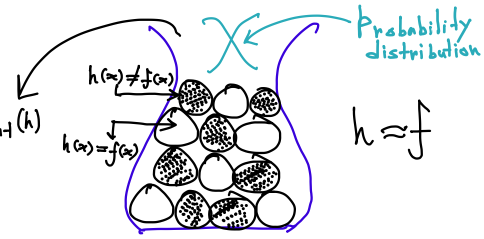
N

Estimate $\nu \leftarrow$ Prob(Dark)

Large N
$\nu \sim \mu$ (within ϵ)

$\mathbb{P}[|\nu-\mu| > \epsilon] \leq 2e^{-2\epsilon^2 N}$

Hoeffding Ineq



$h \approx f$

$E_{in}(h)$

$P[|E_{out}(h) - E_{in}(h)| > \epsilon] \leq 2e^{-2\epsilon^2 N}$

h_1 h_2 h_3 M

$P[|E_{out}(g) - E_{in}(g)|] \leq 2Me^{-2\epsilon^2 N}$

P.A.C

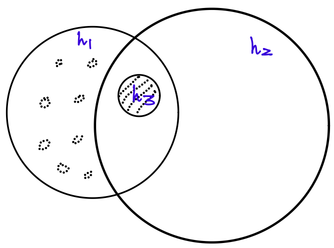
Concept Learning

Boolean Formula $\Rightarrow$ Anded

h_1: <Sunny ? ? Strong ? ?> $\rightarrow$ (Sky = Sunny) $\wedge$ (Wind = strong)

h_2: <Sunny ? ? ? ? Same> $\rightarrow$ $\begin{pmatrix} S = \text{Sunny} \\ \wedge\ F = S \\ \wedge\ \phi \end{pmatrix} = F$

h_3: <Sunny ? ? Strong ? Same>



$< \phi\ \phi\ \phi\ \phi\ \phi\ \phi >$
-1
$< ?\ ?\ ?\ ?\ ?\ ? >$

Sky = 3 (1)
oth = 2 (5)
$M = 1 + 4 \times 3^5$

Find S

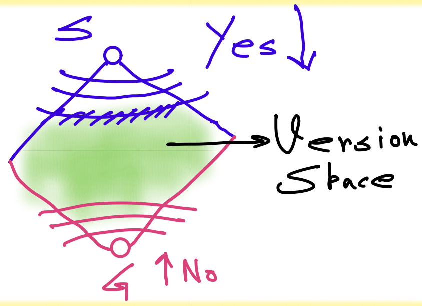
$S_0 \leftarrow <\phi, \phi, \phi, \phi, \phi, \phi>$

$S_1 \leftarrow$ < Sunny, W, N, strong, W, Same>

$S_2 \leftarrow$ < Sun, W, ?, str, W, S>

S_3

S_4: < Sun, W, ?, str, ?, ?>



Candidate Elimination

$\log_2 |VS|$

$S_0 \leftarrow \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

$S_1 \leftarrow \langle S, W, N, S, W, S \rangle$

$S_2 \leftarrow \langle S, W, ?, S, W, S \rangle$

$S_3 = S_2$

$S_4 \rightarrow \langle S, W, ?, S, ?, ? \rangle$

$G_0 \leftarrow \langle ?, ?, ?, ?, ?, ? \rangle$

$G_1 = G_2$

$G_3 \leftarrow$

$\langle S, ?, ?, ?, ?, ? \rangle \langle ?, W, ?, ?, ?, ? \rangle \langle ?, ?, ?, ?, ?, C \rangle$

$G_4 \leftarrow \langle S, ?, ?, ?, ?, ? \rangle \langle ?, W, ?, ?, ?, ? \rangle$

$\langle S, W, ?, ?, ?, ? \rangle \quad \langle ?, W, ?, S, ?, ? \rangle \quad \langle S, ?, ?, ?, S, ?, ? \rangle$

Mitchell's Chapter 2

Sun	Y
Cloudy	Y
Rain	N

OR

	Season	Weather	Play Football
Train	Winter	Sunny	Y
	Summer	Rainy	Y
	Winter	Rainy	N
	Summer	Sunny	?

$S_0 \rightarrow \langle \phi, \phi \rangle$

$G_0 \rightarrow \langle ?, ? \rangle$

T_1:

$S_1 \rightarrow \langle W, S \rangle$

$G_1 \rightarrow \langle ?, ? \rangle$

T_2:

$S_2 \rightarrow \langle ?, ? \rangle$

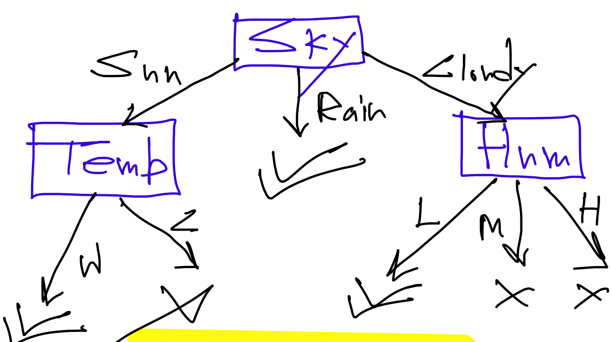
$G_2 \rightarrow \langle ?, ? \rangle$

T_3: – X

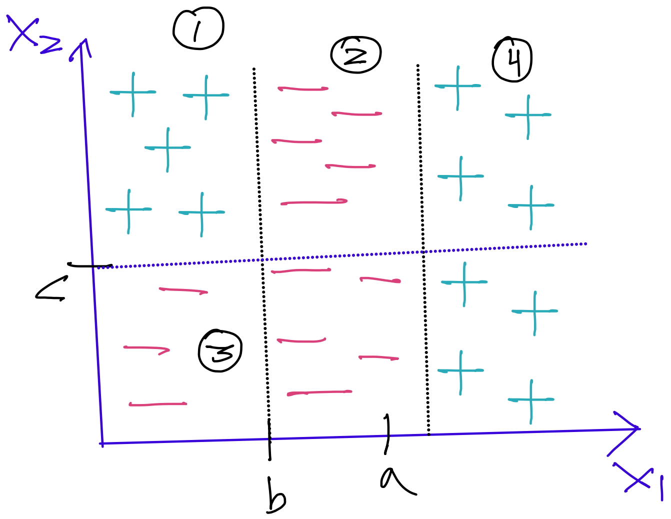
OR $(Sky = Sunny) \wedge (Temp = Warm)$

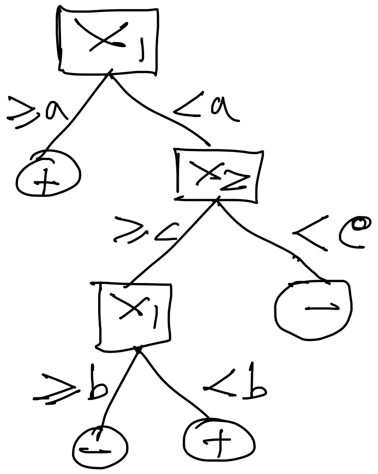
$\vee (Sky = Rainy)$

$\vee (Sky = Cloudy) \wedge (Hum = Low)$

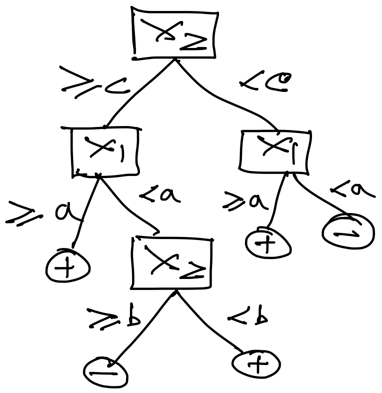


Decision Tree





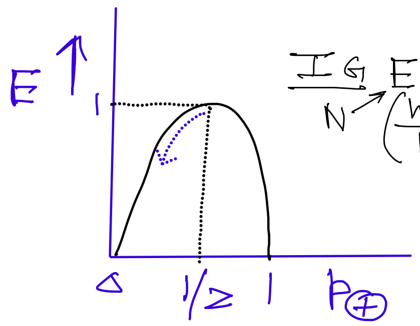
$$\frac{1+3+3+2}{4} = 2.25$$



$$\frac{2+3+3+2+2}{4} = 2.4$$

ID3
1987
Quinlan

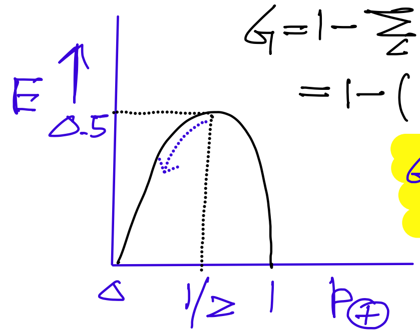
Entropy$(p_\oplus, p_\ominus)$
$$= -p_\oplus \log_2 p_\oplus - p_\ominus \log_2 p_\ominus$$

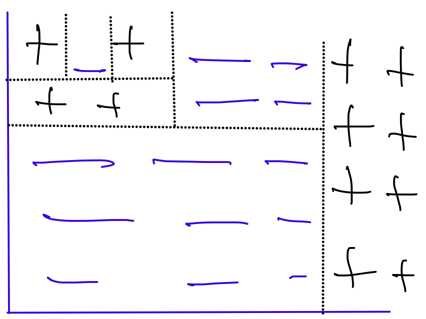


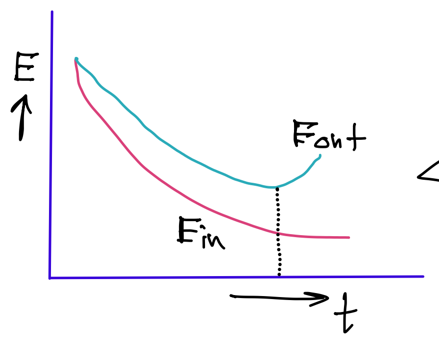
$$IG = E - \left(\frac{n_1}{N}E_1 + \frac{n_2}{N}E_2\right)$$

$$G = 1 - \sum_i p_i^2$$
$$= 1 - (p_\oplus^2 + p_\ominus^2)$$

Gini Index
"Impurity"

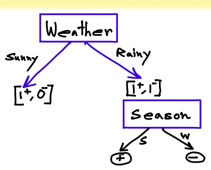


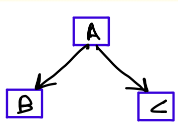




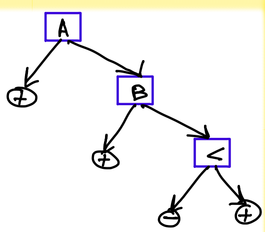
Overfitting

Season	Weather	Play Football
Winter	Sunny	Y
Summer	Rainy	Y
Winter	Rainy	N
Summer	Sunny	?

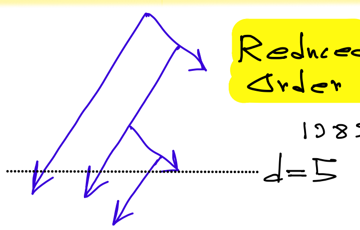




$>$



Occam's Razor



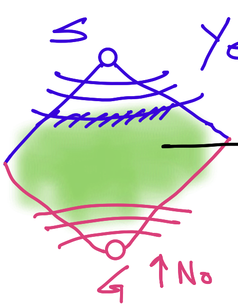
Reduced Order Pruning
1989

d=5

First construct
Full Tree and
Then Eliminate

$(a \wedge b \wedge \cancel{x} \wedge d \wedge \cancel{e})$
$\vee (a \wedge e)$
$\vee (a \wedge \cancel{f} \wedge \cancel{b} \wedge c \wedge d)$

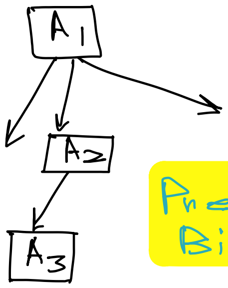
Rule Post Pruning



Version Space

Candidate Elimination

Restriction Bias



ID3

Preference Bias

Temp	10	12	14	22	30	38	42	48
Play	N	N	N	N	Y	Y	N	N

$T \geq 26.5$ $\quad$ $T \leq 40$

	A_1	A_2	A_3	Out
	a_{11}	a_{21}	a_{31}	Y
	a_{12}	a_{22}	a_{32}	Y
DT_1 0	a_{13}	a_{23}	a_{33}	N
	$\square$	a_{24}	a_{34}	N
1 DT_2	a_{15}	a_{25}	a_{35}	Y

$$f: x \rightarrow y \rightarrow \begin{cases} +1 \\ -1 \\ \vdots \end{cases}$$

$$g(x) = ?$$

$$P(y=1 \mid x_{new}) = ?$$

$$P(A=male) = \frac{|S_{male}|}{|S|}$$

$$0 \leq P(A) \leq 1$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \mid B)\, P(B) = P(A \cap B)$$

$$= P(B \mid A)\, P(A)$$

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Prior

Marginal

$$P(A_k \cap B) = \frac{P(B \mid A_k)\, P(A_k)}{\sum_i P(B \mid A_i)\, P(A_i)}$$

$2^N \rightarrow$ Represent JPDT

$\rightarrow 2^{100}$ Rows $\approx 10^{30}$ "Smart" Estimate Probability

$0 \leftarrow 0.00\%$ $\quad 10^{3}$

① 51 Heads + 49 Tails

$$\hat{\theta}_H = \hat{P}(x=1) = 0.51$$

② 2 Head + 1 Tails

$$\hat{\theta}_H = \hat{P}(x=1) = 0.67$$

③ α_H Heads + α_T Tails (Online)

20

$$\frac{\alpha_H + 10}{(\alpha_H + 10) + (\alpha_T + 10)}$$

H H T T H

$\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$

$\hat{\theta}$

θ — $1/2$, $2/5$, $3/5$

$\underset{\theta}{\operatorname{argmax}} \, \mathbb{P}[\text{Data} \mid \theta] \rightarrow \hat{\theta}$

MLE

$$\frac{\partial}{\partial \theta}\left[\mathbb{P}(D \mid \theta)\right] = 0$$

$$\frac{\partial}{\partial \theta}\left[\alpha_H \ln(\theta) + \alpha_T \ln(1-\theta)\right] = 0$$

$$\Rightarrow \frac{\alpha_H}{\theta} - \frac{\alpha_T}{1-\theta} = 0 \Rightarrow \theta = \frac{\alpha_H}{\alpha_H + \alpha_T}$$

MAP

$$\underset{\theta}{\operatorname{argmax}} \left[P(D \mid \theta) P(\theta)\right]$$

Marginal "Independence"

$$P(x_1, x_2) = P(x_1) P(x_2) \qquad x_1 \underset{\text{ind}}{\sim} x_2$$

Cond "Independence"

$$P(x_1 \mid x_2, y) = P(x_1 \mid y)$$

or, $$P(x_1 x_2 \mid y) = P(x_1 \mid y) \cdot P(x_2 \mid y)$$

$$P(x_1, \dots, x_N \mid y) = \prod_i P(x_i \mid y)$$

(1) $$\frac{P(y=y_k) \prod_i P(x_i \mid y=y_k)}{\sum_j P(y=y_j) \prod_i P(x_i \mid y=y_j)} \geq N$$

Naive Bayes

$$\Rightarrow \frac{P[y=1 \mid x_1, \dots, x_N]}{P[y=0 \mid x_1, \dots, x_N]} \gtrless 1$$

$$\Rightarrow \frac{P(y=1) \prod_i P(x_i \mid y=1)}{P(y=0) \prod_i P(x_i \mid y=0)} \gtrless 1$$

Log Linear Model

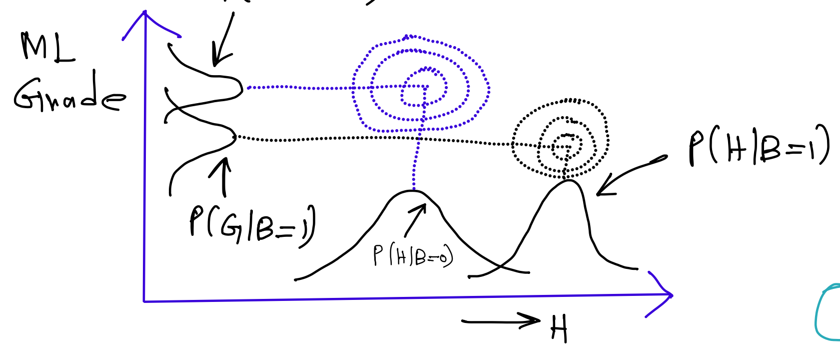
$\hat{\theta}_{i1}$

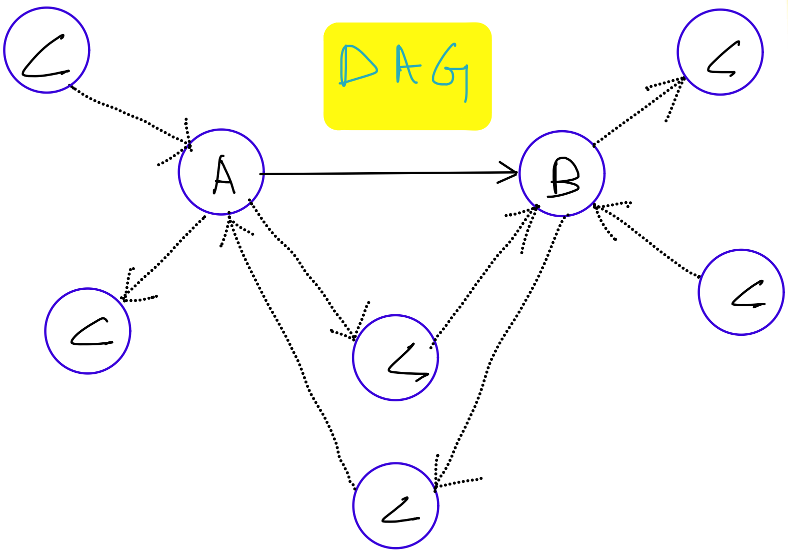
$$\Rightarrow \log\left(\frac{P(y=1)}{P(y=0)}\right) + \sum_i \log\left(\frac{P(x_i \mid y=1)}{P(x_i \mid y=0)}\right) \gtrless 0$$

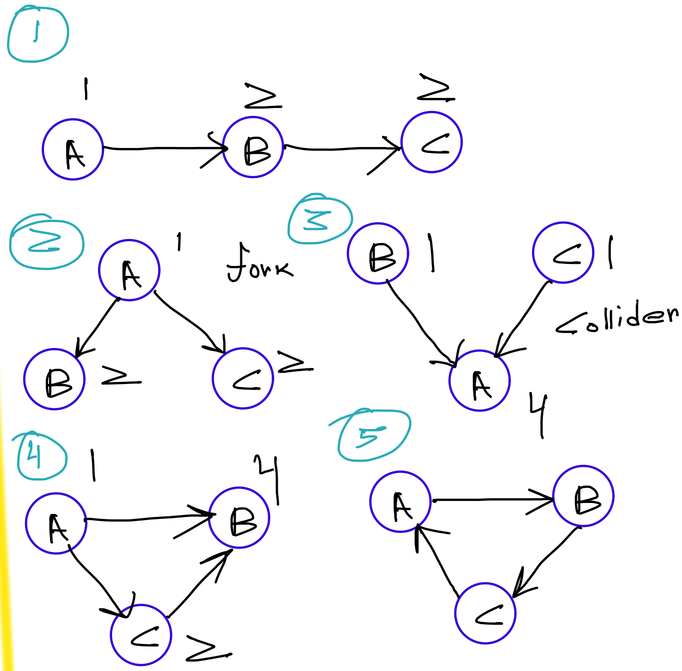
linear wrt x_i $\qquad \hat{\theta}_{i0}$

$$\Rightarrow \log\left(\frac{\hat{\theta}_1}{\hat{\theta}_0}\right) + \sum_i \left(x_i \log\left(\frac{\hat{\theta}_{i1}}{\hat{\theta}_{i0}}\right) + (1-x_i) \log\left(\frac{1-\hat{\theta}_{i1}}{1-\hat{\theta}_{i0}}\right)\right) \gtrless 0$$

$P[\text{Play}=B \mid \text{Height}, \text{ML Grade}]$



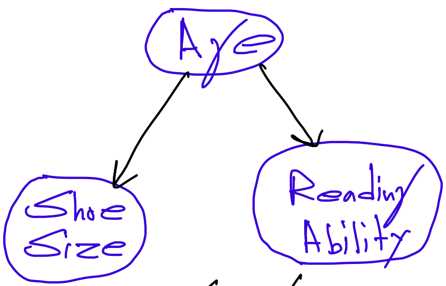


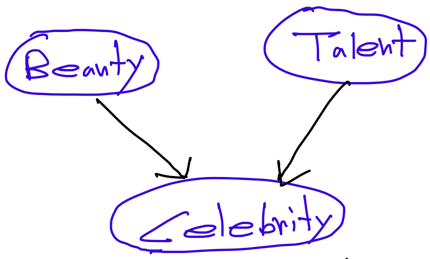


Fire → Smoke → Alarm $F \perp A \mid S$

$$P(FSA) = \cancel{P(A|SF)}\ P(S|F)\ P(F)$$

$$P(A|S)$$





$S \not\perp R \mid \emptyset$

$S \perp R \mid \emptyset$

$B \perp T \mid \emptyset$

$B \not\perp T \mid C$

$G = (V, E)$

$$P(x_1, \dots, x_n) = \prod_i P(x_i \mid Par(x_i))$$

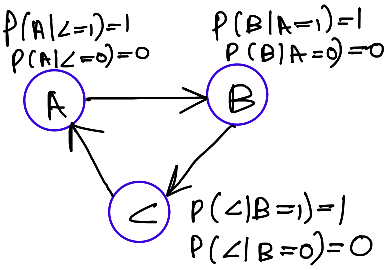
1

Given:- $P(ABC) = P(C|B)\,P(B|A)\,P(A)$

$$P(AC|B) = \frac{P(ABC)}{P(B)} = \frac{P(C|B)\,P(B|A)\,P(A)}{P(B)} = P(C|B)\,P(A|B)$$

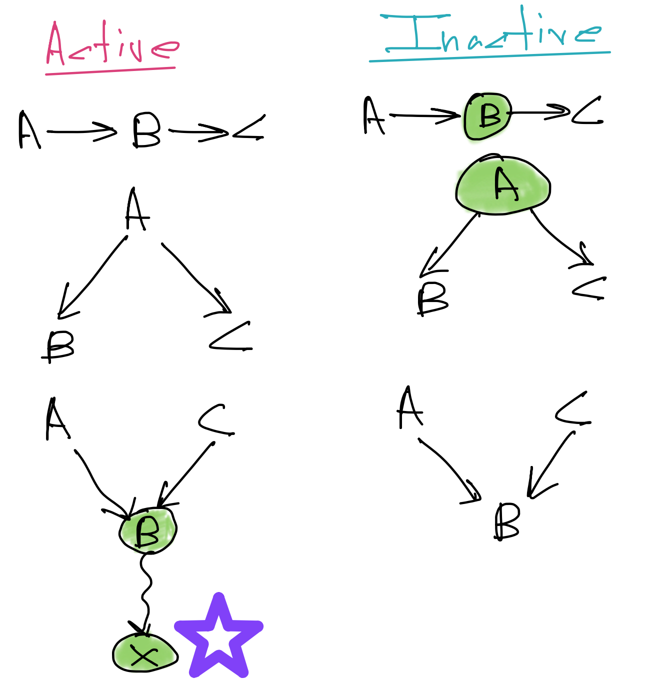
(2) $P(BC|A)$

$$= \frac{P(ABC)}{P(A)} = \frac{P(B|A)\,P(C|A)\,P(A)}{P(A)}$$



$P(ABC)=1$

$P(A=0, B=0, C=0) = 1$



$$L - \underbrace{S - \overbrace{R - W}^{\text{Active}}}_{}$$

$L - S$ Active

$$L - \underbrace{P - \overbrace{R - W}^{\text{Active}}}_{\text{Active}}$$

Non-Parametric Learning:-

↳ Instance / Case Based

K- Nearest Neighbour:-

distance=?

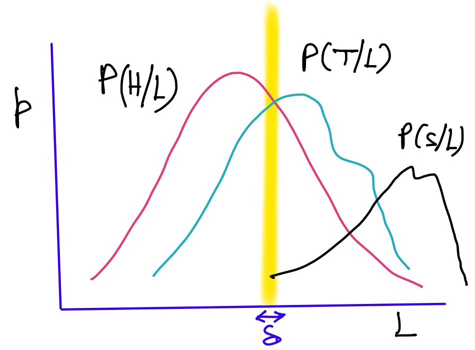
K=?

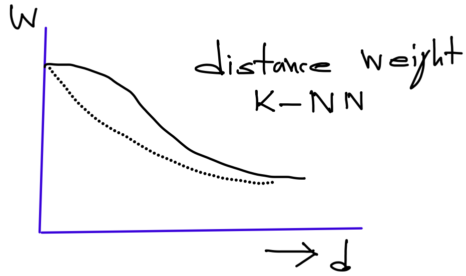
$\mathcal{L} = \{+, -, 0, *\}$

$TE \rightarrow (x_1, l_1) \ldots\ldots\ldots (x_N, l_N)$

$x \longrightarrow (x_{i1}, \ldots, x_{iK})$

$$\arg\max_{l} \sum_{j=1}^{K} \delta(l, f(x_{ij})) \cdot w_j(d(x, x_{ij}))$$

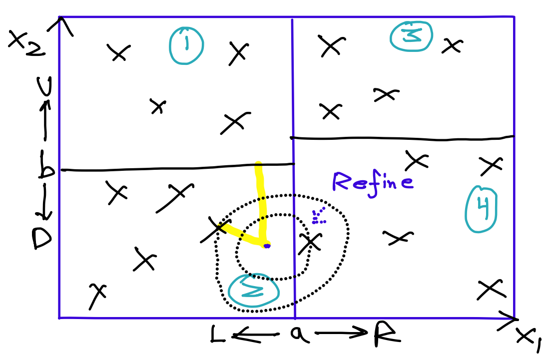


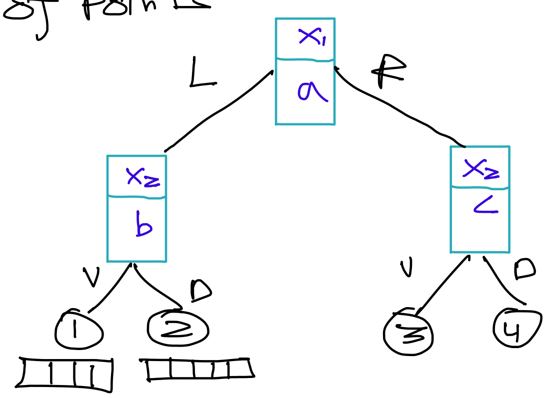


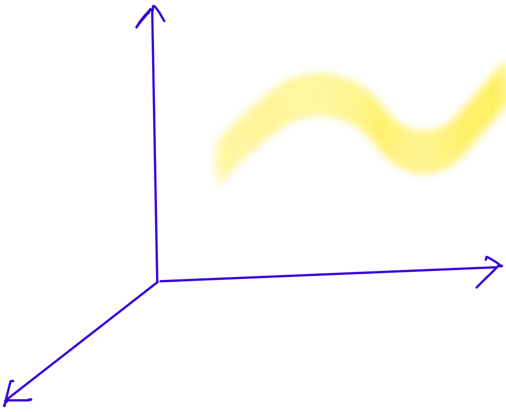
1. $O(Nd) \longrightarrow$ distance calc
2. $O(KN) / O(K + (N-K)\log K) \leftarrow$ Max Heap
3. $O(K) \leftarrow$ Bubble Sort

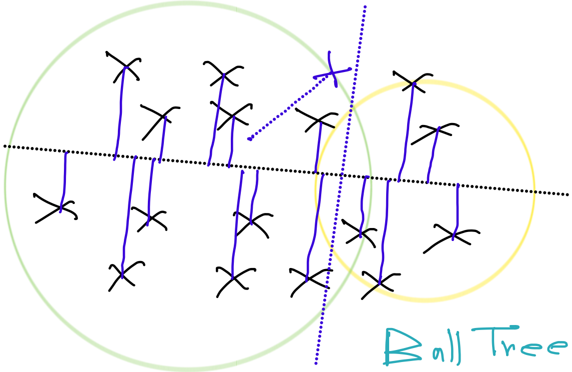
Condensing:- Minimal consistent set of points

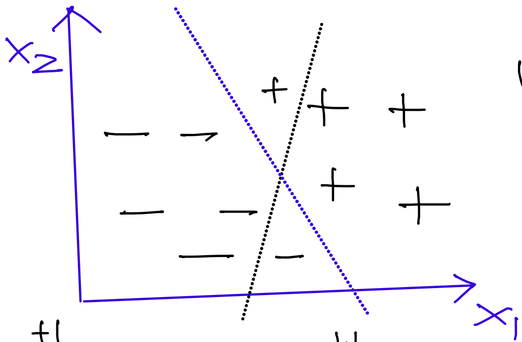
K-D Tree:-











$$w_1x_1+w_2x_2+w_0x_0=0$$

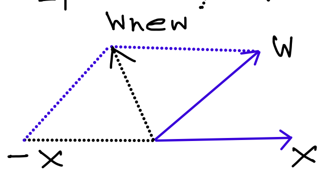
$$\text{sign}\left(\sum_{i=0}^{d} w_i x_i\right) \longrightarrow \text{sign}(w^T x)$$

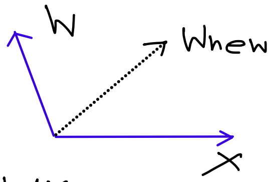
$$W=\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \qquad X=\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$$

+1   +1

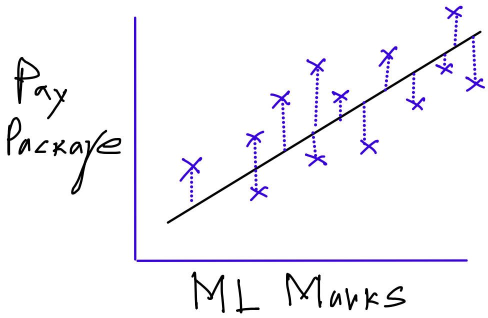
$$w \cdot x_i \neq y_i$$

-1   -1





$$W_{new} \leftarrow W + yx$$



$$E_{in}=\frac{1}{N}\sum_{n=1}^{N}\left(w^T x_n - y_n\right)^2$$

$$\frac{\partial E_{in}}{\partial w}=0=\frac{1}{N}\|Xw-y\|^2$$

$$X=\begin{bmatrix} \text{—} & x_1^T & \text{—} \\ & & \\ \text{—} & x_N^T & \text{—} \end{bmatrix}_{N\times(d+1)} \qquad W=\begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}_{(d+1)\times 1} \qquad Y=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{N\times 1}$$

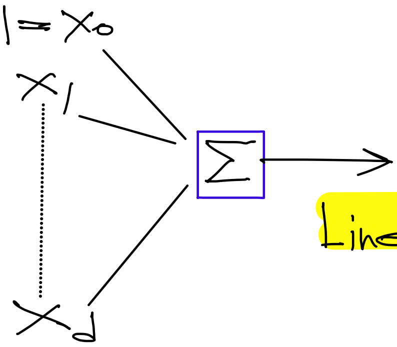
$$\frac{\partial E_{in}}{\partial w}=\frac{2}{N}X^T(Xw-y)=0$$

$$w=\underbrace{(X^TX)^{-1}X^T}_{\text{Psendo-Inverse}}\,y$$

$$\underbrace{[(d+1)\times N]\;[N\times(d+1)]}\quad$$

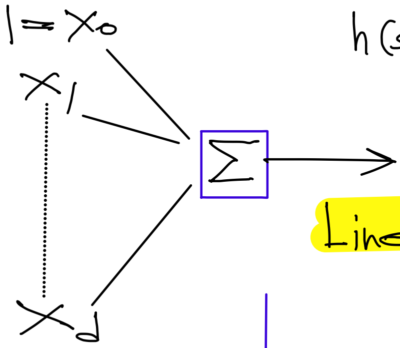
$$[(d+1)\times(d+1)]\;[(d+1)\times N]$$

$$\Longrightarrow [(d+1)\times N]\;[N\times 1]$$



$$s = w^T x = h(x)$$

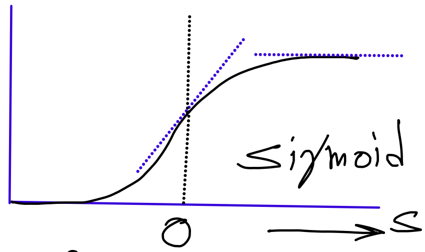
Linear Regression



$h(s) = \text{sign}(s)$

$$s = w^T x$$

Linear Classification



$$P[y \mid x] = \begin{cases} h(x) & x > 0 \\ 1 - h(-x) & x < 0 \end{cases}$$

$$h(s) = \frac{e^s}{1+e^s} = \frac{1}{1+e^{-s}}$$

$$\tanh(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}$$

$$h(s) = h(y w^T x)$$

$h(-s) = 1 - h(s)$

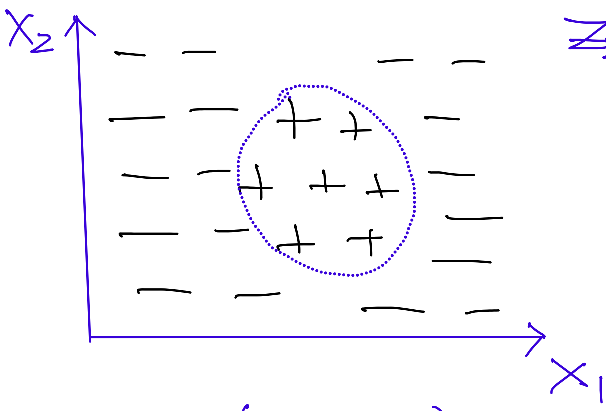
$$\underset{w}{\operatorname{argmax}} \prod_{i=1}^{N} h(y_i w^T x_i)$$

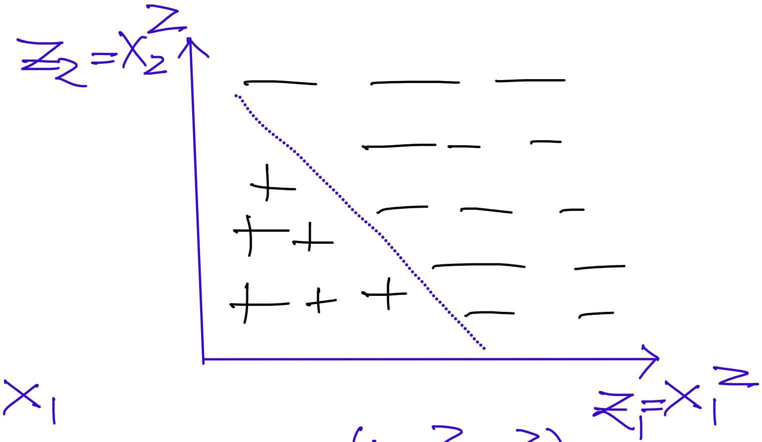
$$\operatorname{Max} \prod_{i=1}^{N} \frac{1}{1+e^{-y_i w^T x_i}} \equiv \operatorname{Min} \frac{1}{N} \sum_{i=1}^{N} \underbrace{\ln\left(1+e^{-y_i w^T x_i}\right)}_{\text{Cross Entropy}}$$

$$x_i y_i \qquad \frac{\partial E_{in}}{\partial w} = 0$$

Learning from Data

$$\nabla E_{in} \quad \frac{1}{N} \sum_{i=1}^{N} \frac{-y_i x_i}{1+e^{y_i w^T x_i}} \qquad w' \leftarrow w + \eta$$





$$x = (1, x_1, x_2) \xrightarrow[\quad z = \phi(x) \quad]{} (1, x_1^2, x_2^2)$$

Data Snooping

$$w = \phi^{-1}(\tilde{w}) \qquad \tilde{w}$$

$$(1, x_1, x_2, x_1^2, x_2^2, x_1 x_2) \Longrightarrow (1, x_1^2 + x_2^2)$$

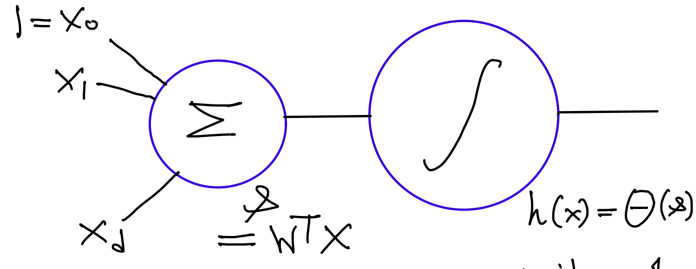
$(x_1,+1),(x_2,0)\ldots\ldots(x_N,+1)$ $(x_{new}, ?\%)$

$$\mathbb{P}[y|x] = \begin{cases} h(x), & \text{when } y=1 \\ 1-h(x), & \text{when } y=0 \end{cases}$$

$$= \begin{cases} \Theta(s) & y=+1 \\ \Theta(-s) & y=-1 \end{cases}$$

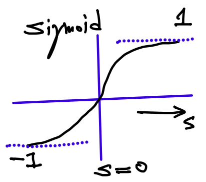
$$= \Theta(ys)$$

$$= \frac{1}{1+e^{-yw^Tx}}$$



$$\Theta(s) = \frac{1}{1+e^{-s}}$$

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



max

$$\prod_{i=1}^{N} \hat{\mathbb{P}}(y_i|x_i) = \prod_{i=1}^{N} \frac{1}{1+e^{-y_iw^Tx_i}} \equiv \sum_{i=1}^{N} \ln\left(\frac{1}{1+e^{y_iw^Tx_i}}\right)$$

min

$$E = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\ln\left(1+e^{-y_iw^Tx_i}\right)}_{\text{Cross Entropy Error}}$$

Batch GD

$$\frac{\partial E}{\partial w} = 0$$

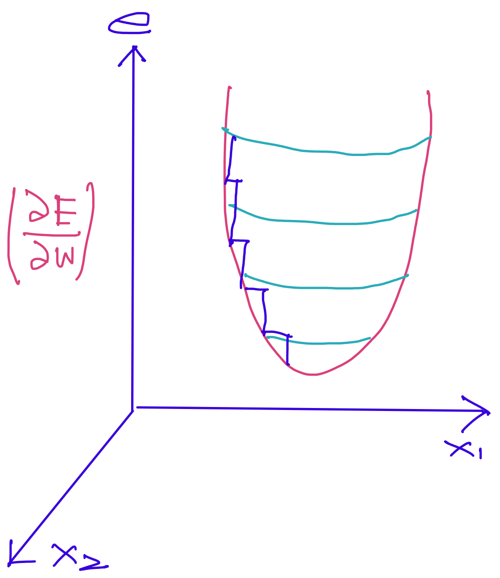
$$\frac{\partial E}{\partial w} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1+e^{-y_iw^Tx_i}} e^{-y_iw^Tx_i}(-y_ix_i)$$

For each (x_i, y_i)

$$\frac{\partial E}{\partial w} = \frac{-y_ix_i}{1+e^{y_iw^Tx_i}}$$

Stochastic Gradient Descent

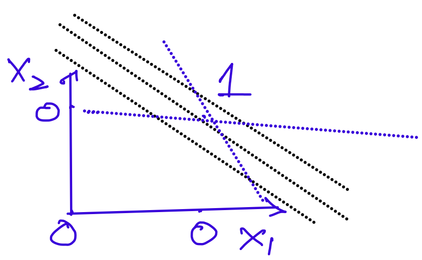
$$w' \leftarrow w + \eta\left(\frac{\partial E}{\partial w}\right)$$



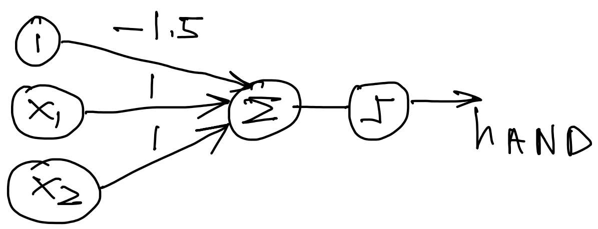
Logistic Regression

Ex:- Credit Repay %

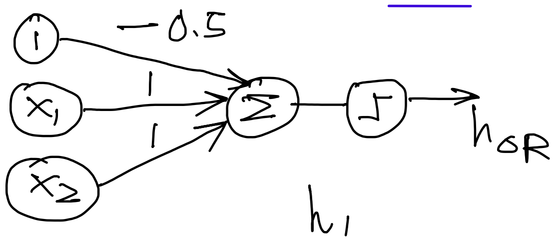
Rainfall Chance %



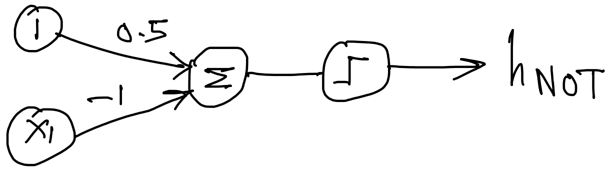
AND

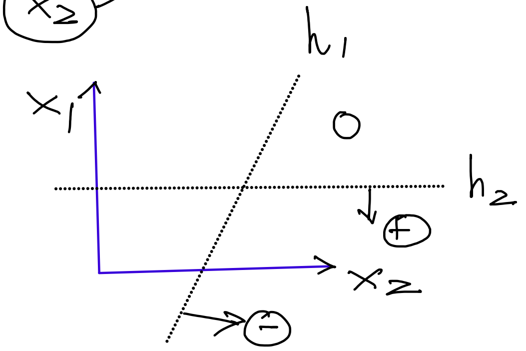


OR

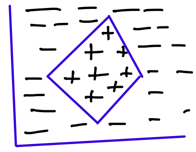


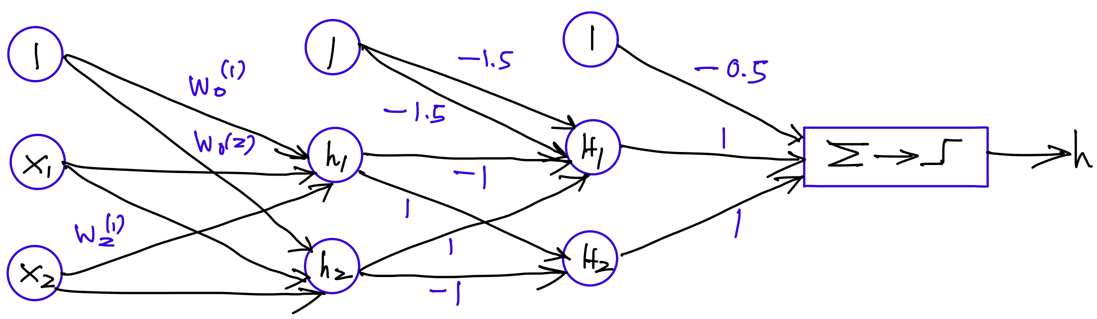
NOT





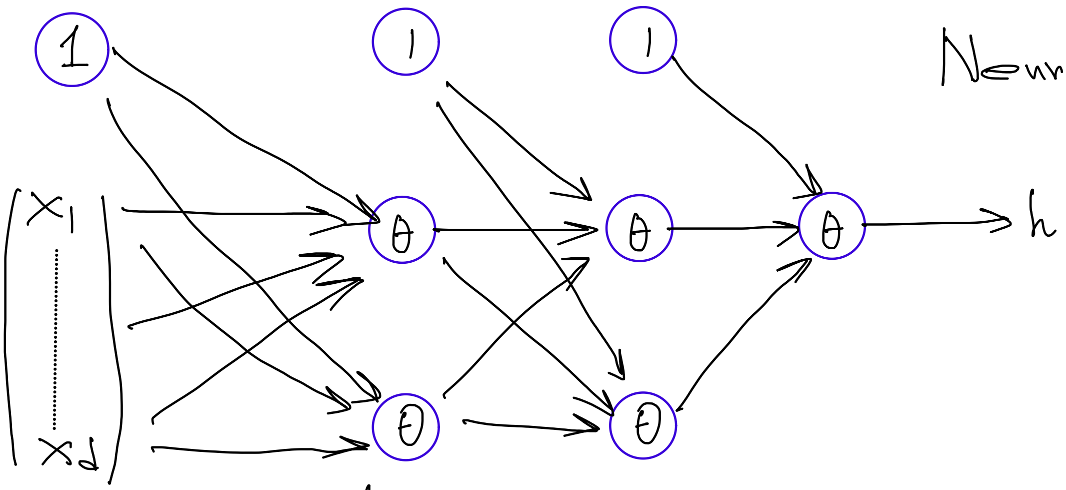
$$h_{XOR} = \underbrace{(\hat{h}_1 \wedge h_2)}_{H_1} \vee \underbrace{(h_1 \wedge \hat{h}_2)}_{H_2}$$



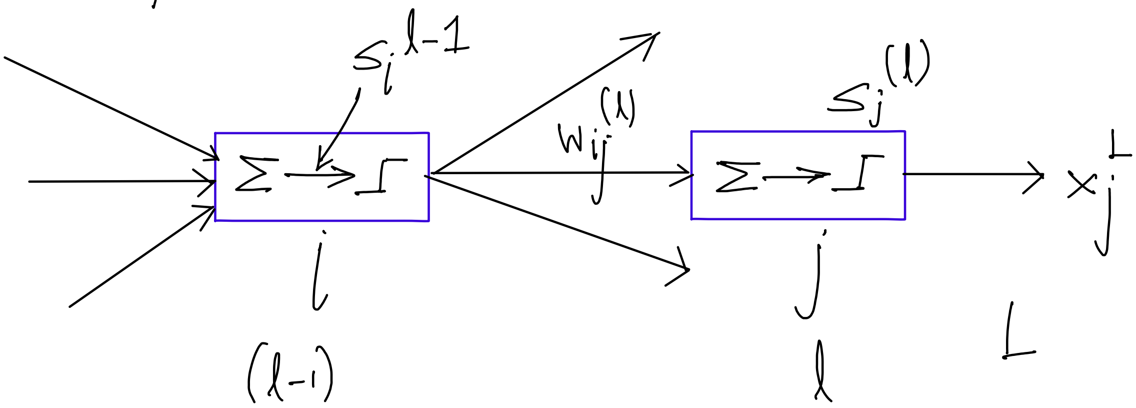


ANN

Multi Layer Perception Network (Feed-Fwd)



Neural Network



$$w_{ij}^{(l)} = \begin{cases} 1 \le l \le L & \text{layers} \\ 0 \le i \le d^{(l-1)} & \text{input} \\ 1 \le j \le d^{(l)} & \text{output} \end{cases}$$

$$(x_0^{(0)}=1, x_1^{(0)} \dots x_d^{(0)})$$

$$x_j^{(l)} = \theta(s_j^{(l)}) = \Theta\left(\sum_{i=0}^{d} w_{ij}^{(l)} x_i^{(l-1)}\right)$$

$$\Theta(s) = \tanh(s) = \frac{e^s - e^{-s}}{e^s - e^{-s}}$$

$$x^0 \longrightarrow x^{(1)} \longrightarrow x^{(2)} \cdots\cdots \longrightarrow x_1^{(L)}$$

$$\frac{\partial e}{\partial w_{ij}^{(l)}} = \frac{\partial e}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)}$$

$$w_{ij}^{(l)'} \leftarrow w_{ij}^{(l)} + \eta \frac{\partial e}{\partial w_{ij}^{(l)}}$$

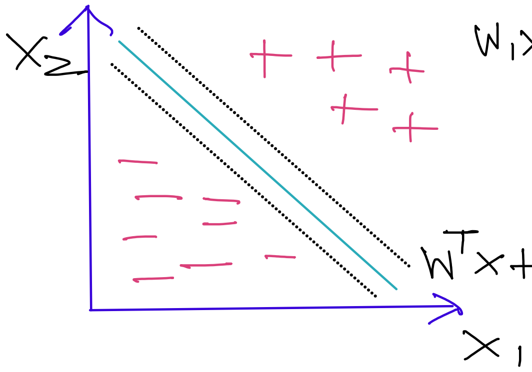
$$\delta_1^{(L)} = \frac{\partial e}{\partial s_1^{(L)}} = \frac{\partial}{\partial s_1^{(L)}}(\theta(s_1^L) - y)^2$$

$$\delta_i^{(l-1)} = \frac{\partial e}{\partial s_i^{(l-1)}}$$

$$= \sum_{j=1}^{d^{(l-1)}} \frac{\partial e}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}} * \frac{\partial x_i^{(l-1)}}{\partial s_i^{(l-1)}}$$

$$= \delta_j^{(l)} * w_{ij}^{(l)} * \theta'(s_i^{l-1})$$

Back Propagation



$$w_1x_1 + w_2x_2 + b = 0$$

$$d_i = \frac{|w_1x_1 + w_2x_2 + b|}{\sqrt{w_1^2 + w_2^2}}$$

Constraint $y_i(w^Tx_i+b) \geq 0 \;\; \forall i$

$$\max_{w,b}\left[\min_i (d_i) = \frac{|w^Tx+b|}{||w||}\right]$$

Constraint:- $y_i(w^Tx_i+b) \geq 1$ for all i opt $\min \frac{1}{2} w^Tw$ s.t

Primal Opt Problem

Dual OPT Prob:-   KKT

$$\min \frac{1}{2} w^Tw + \sum_{i=1}^{N} \alpha_i(1 - y_i(w^Tx_i+b))$$

$$\text{s.t } \alpha_i \geq 0$$

$$\alpha_i = 0 \Rightarrow y_i(w^Tx_i+b) > 1$$

$$\alpha_i > 0 \Rightarrow y_i(w^Tx_i+b) = 1$$

$$\frac{\partial L}{\partial W} = 0 \Rightarrow W - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \Longrightarrow W = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$L = \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i y_i y_j (x_i \cdot x_j) \alpha_j + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i y_i y_j (x_i x_j) \alpha_j$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i y_i y_j (x_i \cdot x_j) \alpha_j$$

$$= \Lambda^T V - \frac{1}{2} \Lambda^T H \Lambda$$

$$H = \begin{bmatrix} & \vdots & \\ \cdots & y_i y_j x_i x_j & \cdots \\ & \vdots & \end{bmatrix}$$

"Hessian"

$$\Lambda = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$$

$$W^T X_{sv} + b = 1$$

Support Vector Machine

$$W = \sum_{i \in SV} \alpha_i y_i X_i$$

Const

Issues:

↳ Noise (1) $\Rightarrow$ Slack ξ_i $\quad y_i(W^T x_i + b) \geq 1 - \xi_i \quad (\alpha_i)$

$$\mu_i \rightarrow \xi_i \geq 0$$

$$\text{s.t} \quad \min_{w,b} \left(\frac{1}{2} W^T W + C \sum \xi_i \right)$$

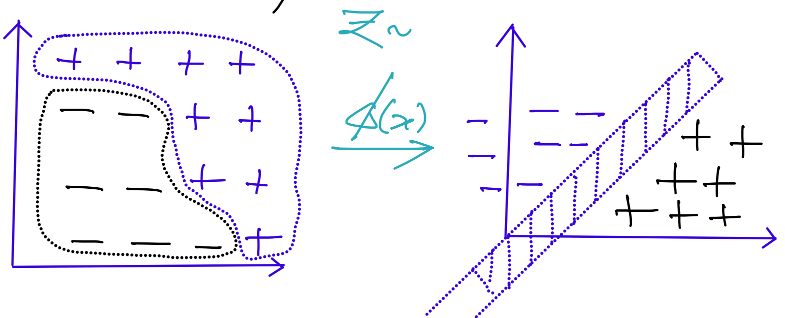
Dual:
$$\min_{w,b} \frac{1}{2} W^T W + \sum_{i=1}^{N} \alpha_i \left(1 - \xi_i - y_i (W^T x_i + b)\right) - \sum_{i=1}^{N} \mu_i \xi_i$$

$$\text{s.t} \quad \alpha_i \geq 0, \ \mu_i \geq 0$$

$$\frac{\partial L}{\partial W} = 0 \qquad \frac{\partial L}{\partial b} = 0 \qquad \frac{\partial L}{\partial \xi_i} = 0$$

$$W = \sum \alpha_i y_i x_i \qquad \sum \alpha_i y_i = 0 \qquad C = \sum \mu_i + \alpha_i$$

Non-Linearity:



$$X=(1,x_1,x_2) \xrightarrow{\phi} Z=(1,x_1,x_2,x_1^2,x_2^2,x_1x_2)$$
$$X'=(1,x_1',x_2') \xrightarrow{\phi} Z'=(1,x_1',x_2',x_1'^2,x_2'^2,x_1'x_2')$$

$$K(X,X') = X\cdot X' \quad \text{linear}$$
$$= (1+X\cdot X')^2$$
$$= 1+2x_1x_1'+2x_2x_2'+x_1^2x_1'^2+x_2^2x_2'^2$$
$$+2x_1x_1'x_2x_2'$$
$$= (1,\sqrt{2}x_1,\sqrt{2}x_2,x_1^2,\sqrt{2}x_1x_2)^T$$
$$(1,\sqrt{2}x_1',\sqrt{2}x_2',x_1'^2,\sqrt{2}x_1'x_2')^T$$

"Kernel"

Gaussian Kernel

$\boxed{\text{Ref} - [2],[6]}$

Classifier Evaluation :-

Perf Metric | Eval Method | Compare Model

		Predicted Class: Yes	Predicted Class: No
Actual Class	Yes	TP	FN
Actual Class	No	FP	TN

Confusion Matrix

$$\text{Accuracy } (A) = \frac{|TP|+|TN|}{|TP|+|TN|+|FP|+|FN|}$$

$$\text{Recall} = \frac{|TP|}{|TP|+|FN|}$$

$$\text{Precision} = \frac{|TP|}{|TP|+|FP|}$$

F-Score (F_1 — Measure)

$$\frac{1}{F} = \frac{1}{2}\left(\frac{1}{P}+\frac{1}{R}\right) \Rightarrow F = \frac{2PR}{P+R}$$

$$W.A = \frac{w_1|TP|+w_2|TN|}{w_1|TP|+w_2|TN|+w_3|FP|+w_4|FN|}$$

Eval Method :- ① Holdout $\frac{2}{3}$ Train + $\frac{1}{3}$ Test

c Train + $(N-c)$ Test

② Random Subsplitting → Random frac of C

K

N

③ Cross Validation

④ Stratefied Sampling

(x1, Y)
(x2, N)
(x3, Y)
(x4, N)
(x5, Y)
(x6, Y)

X1 X3
X5 X6
Y

X2 X4
N

Bootstrapping
(Replace)

ROC (Receiver Operating Chair)