ML-Final-21. Aug. 2017

$\boxed{Q1}$ 1- Naive Bayes: $2(n+1)$ or $2n+1$
logistic Regression: $n+1$ Bias $P(No) = 1 - P(Yes)$

2- # of hidden nodes
→ More than enough ~ Complexity ~ Variance ~ overfitting
→ less than enough ~ Simplicity ~ Bias ~ Underfitting.

3- Sigmoid

4- True, logistic regression gets better as the training set grows!

5- False

6- True

7- True, if the classes are separable.

8- Decrease / Increase

$\boxed{9-}$ Recognition, Detection / Classification

10 - Choice of Algorithms / Feature Selection / Feature Creation / Model Selection according to Lecture 11

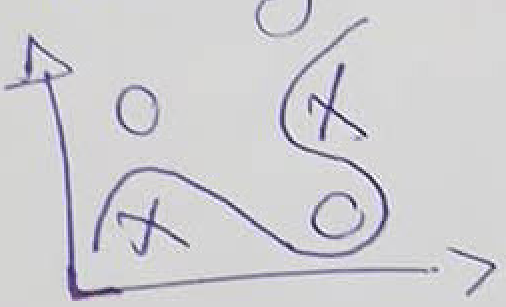
11- Parametric models: - linear regression — why am I writing this?
- logistic regression
- Naive Bayes.

Non-Parametric models: - DTs
- KNN

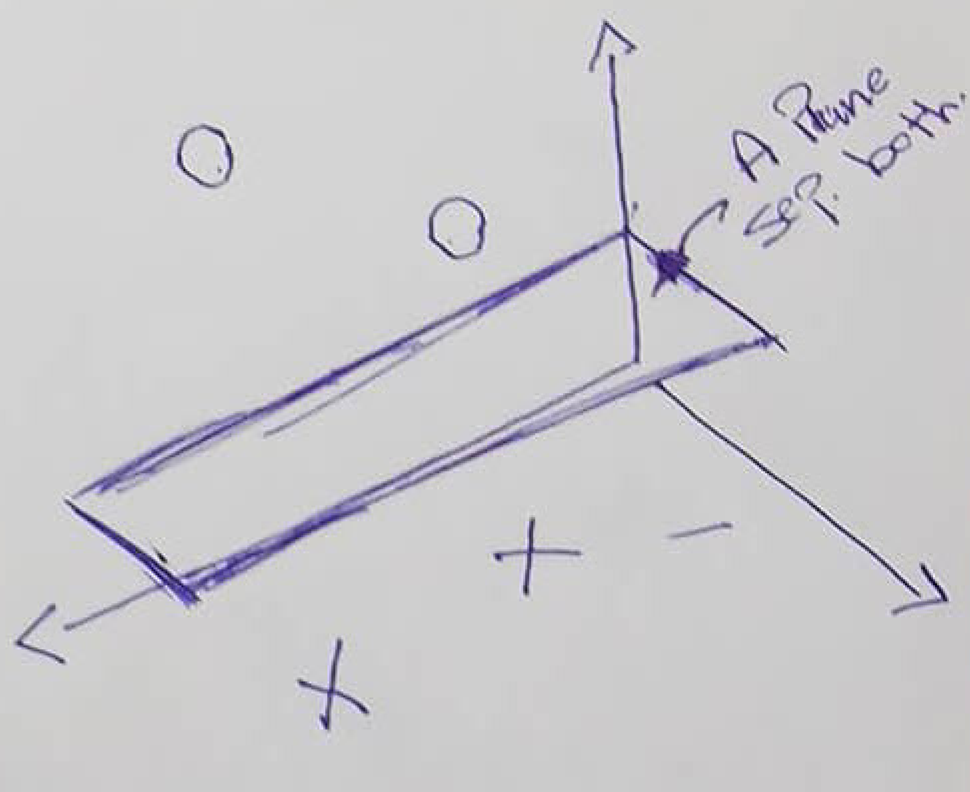
Q2

1) We need to estimate: $P(Y=1)$, $P(X_1=1|Y=0)$
$P(X_2=1|Y=0)$, $P(X_3=1|Y=0)$, $P(X_1=1|Y=1)$,
$P(X_2=1|Y=1)$, $P(X_3=1|Y=1)$, other ~~Params~~ Probs.
Can be obtained with the Constraint that it Sums up to 1.
So we only need to estimate 7 Params.

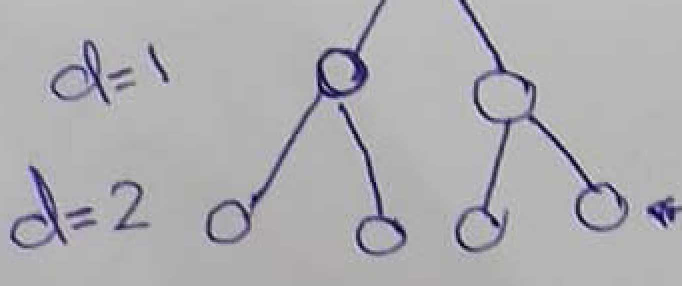
2) (a) Yes, notice that the number of Params to use is not mentioned, and since it's testing on training data, overfitting is not of Concern



(b) SVM? Yes
(Quadratic Kernel)

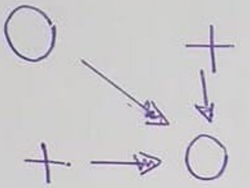


(c) DT-d=2?



← Training data set size = 4
∴ It overfits, memorizes data
∴ Yes!

d) KNN (K = 3)



~ Predicted as + but is O

$\Rightarrow$ No!

3) MSE $= (1/5)\left((-.2)^2 + (-.4)^2 + (-.8)^2 + (1.3)^2 + (-0.7)^2\right)$

$= 0.604$

4) Hierarchical Clustering

* No assumption on the number of Clusters K.
* Reflects taxonomies in data.

K-Means

* Allows for backtracking.
* Relatively efficient.

5) Single Link

Wrong notation, should be {

$$D_{sl}(C_i, C_j) = \min_{x,y} \left(d(x,y) \mid x \in C_i, y \in C_j \right)$$

Complete Link

lol wrong again

$$D_{cl}(C_i, C_j) = \max_{x,y} \left(d(x,y) \mid x \in C_i, y \in C_j \right)$$

b) Relu

* Turns all -ve to 0
* Accounts for non-linear. relationships.

Pooling

* Reduces feature map size, Simplifying Computation in later layers.
* Allows the network to be less sensitive to changes in the location of the feature.

7) True, As we have more and more data, training error increases and testing error decreases.
And they all converge to the true error.

Q3

a) # estimated Params:

Input: $125 * 125 * 3$

feature Maps: $5 * 5 * 10$

not sure if that's what's req.

Output: $\frac{125-5}{3} + 1 = 41$

b) Output: $\frac{125-5+2*3}{3} + 1 = 43$

As $W = H \Rightarrow (43, 43)$

Q4

(a)

1 P_1 P_2 $P_{3,4}$ P_5 P_6 P_7 ~ 1.36

	P_1	P_2	$P_{3,4}$	P_5	P_6	P_7
P_1	0	2.24	3.92	4.55	6.52	8.04
P_2	2.24	0	2.86	2.35	4.53	6.00
$P_{3,4}$	3.92	2.86	0	2.86	2.66	3.91
P_5	4.55	2.35	2.86	0	2.42	3.77
P_6	6.52	4.53	2.66	2.42	0	1.53
P_7	8.04	6.00	3.91	3.77	1.53	0

All this is kind of redundant :

2 P_1 P_2 $P_{3,4}$ P_5 $P_{6,7}$ ~ 1.53

	P_1	P_2	$P_{3,4}$	P_5	$P_{6,7}$
P_1	0				
P_2	$\boxed{2.24}$	0			
$P_{3,4}$	3.92	2.86	0		
P_5	4.55	2.35	2.86	0	
$P_{6,7}$	6.52	4.53	2.66	2.42	0

3 $P_{1,2}$ $\cancel{P_2}$ $P_{3,4}$ P_5 $P_{6,7}$ ~ 2.24

	$P_{1,2}$	$P_{3,4}$	P_5	$P_{6,7}$
$P_{1,2}$	0			
$P_{3,4}$	2.86	0		
P_5	$\boxed{2.35}$	2.86	0	
$P_{6,7}$	4.53	2.66	2.42	0

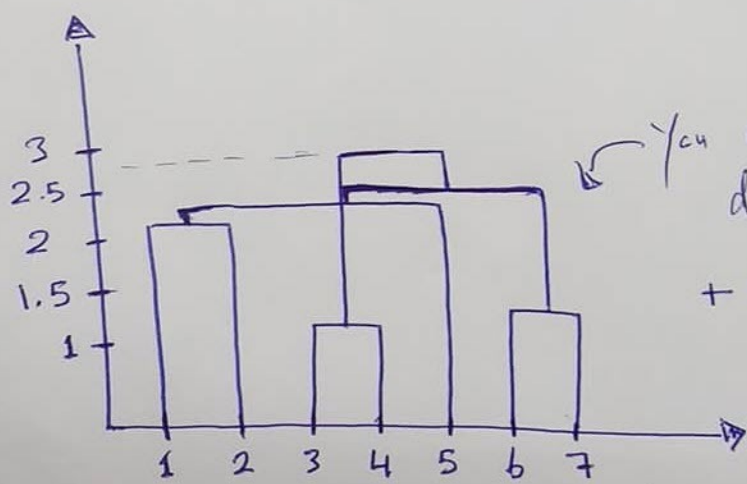
4 $P_{((1,2),5)}$ $P_{3,4}$ $P_{6,7}$ ~ 2.35

$P_{((1,2),5)}$	0		
$P_{3,4}$	2.86	0	
$P_{6,7}$	$\boxed{2.42}$	2.66	0

5 $P_{(((1,2),5),(6,7))}$ $P_{3,4}$ ~ 2.42

$P_{(((1,2),5),(6,7))}$	0	
$P_{3,4}$	$\boxed{2.66}$	0

6 $P_{((((1,2),5),(6,7)),(3,4))}$



You can do better!

+ It looks weird :)

Q5

$$\begin{array}{c} X \\ Y \end{array}\begin{bmatrix} 1.0 & 1.5 & 3.0 & 5.0 & 3.5 & 4.5 & 3.5 \\ 1.0 & 2.0 & 4.0 & 7.0 & 5.0 & 5.0 & 4.5 \end{bmatrix}$$

Iteration 0

D

$$\begin{array}{c} C_1(1,1) \\ C_2(5.0,7.0) \end{array}\begin{bmatrix} 0 & 1.12 & \overset{\sqrt{13}}{3.6} & 7.21 & 4.7 & 5.32 & 4.3 \\ 7.21 & 6.1 & \underset{\sqrt{13}}{3.605} & 0 & 2.5 & 2.06 & 2.92 \end{bmatrix}$$

Assignment

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$C_1\left(\frac{1+1.5}{2}, \frac{1+2}{2}\right) = (1.25, 1.5)$$

~~$C_2(5.026,$~~

$C_2(3.9, 5.1)$

Iteration 1

D

$$\begin{array}{c} C_1(1.25,1.5) \\ C_2(3.9,5.1) \end{array}\begin{bmatrix} 0.56 & 0.56 & 3.05 & 6.67 & 4.16 & 4.78 & 3.75 \\ 5.02 & 3.92 & 1.42 & 2.19 & 0.412 & 0.6 & 0.7 \end{bmatrix}$$

Assignment

$$\begin{array}{c} \text{Cluster 1} \\ \text{Cluster 2} \end{array}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

No movements

Stop!