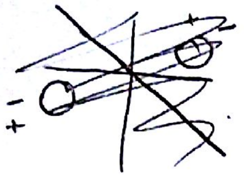
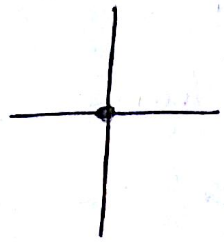
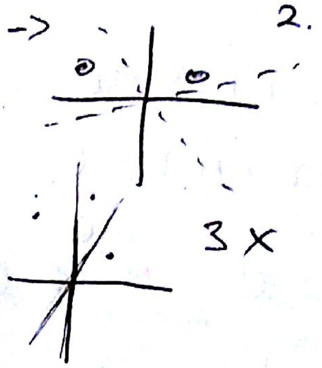
VC.

1) $d=1$ ... → hyperplane = $d=0$

$d=2$ = hyperplane $d-1$







$d=3$

⇒ For any d, ⇒ you have to seperate the points such that each point is by itself at one point.

⇒ Also have to have d_{C_2} ⇒ groups of any 2 points.

⇒ on & on, one have 2^n labels for any set of n-points

Lower bound,

for $d=n$, have ~~any~~ all points where only one coordinante is 1.

i.e. $d=2$ $(0,1)$ $(1,0)$

$d=3$ $(1,0,0)$ $(0,1,0)$ $(0,0,1)$

... → there is always a classifier for any of these points.

Upper-bound,

any set of $d+1$ points ⇒ linearly dependant. ⇒ $x_{d+1} = \sum_{i=1}^{d} \alpha_i \cdot x_i$

Say we have S as set of $n+1$ points. ⇒ we can partition it to 2 sets S_1 & S_2 ⇒ convex hull intersects. ⇒ $p \in S_1$ be the intersection. but with origin

Say there's a hyperplane

$$w \cdot x_1 \le w_0 \quad \forall x_i \in S_1 \quad \Rightarrow \quad w \cdot p \le w_0$$

$$w \cdot x_i > w_0 \quad \forall x_i \in S_2$$

but $\underset{x_{d+1}}{w \cdot p} = \sum_{i:x_i \in S_2} \underset{\alpha_i}{\lambda_i} w \cdot x_i \ge \left(\sum_{i:x_i} \lambda_i\right) \min_{i:x_i \in S} (w \cdot x_i) = \min_{i:x_2 \in S_2} (w \cdot x_i) > w_0$

⇒ contradict

2.

n is the same,

$BIC(M_1) = \ell_1 - \frac{1}{2} p \log n$

$BIC(M_2) = \ell_2 - \frac{1}{2} 10 p \log n$

$M_2 - M_1 \leq \ell_2 - \ell_1 - (9/2 \, p \log n)$

ℓ_2 = higher but not too high to overfit ~~the~~

10x as many ~~datapoints~~ parameters,

$\ell_2 - \ell_1 > 9/2 \, p \log n$

p = originall # of parameters

& n =

3\. a) ~~Since, 4 points & 3 are centers~~

~~→ never going to be a change~~

(-5, 2) (4, 4) (0, -6)

~~(0,0)~~) (0,0)

Cost = ~~[illegible]~~ 6

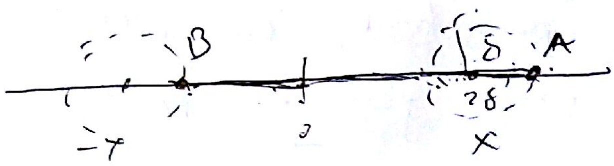
(~~[illegible]~~) (4, 4) (0, ~~[illegible]~~)

~~[illegible]~~ (0, -6)

→ ~~Ca~~ Clusters (-5, 2), (4, 4), {(0, -6), (0,0)}

b) Cluster is {(-5, 2), (0,0)} ,(4, 4) ,(0, -6)

4a) Because $\delta << x$ if we had "randomly" chosen ~~the~~ ~~dae~~ one closest point & one farthest,



2δ

$2x - 2\delta$

$2x - 2\delta > 2\delta$

~~$2x - 2\delta > 2\delta$~~ $\Rightarrow$ no matter what

or $2x - 2\delta < 2\delta$

point we choose, it will end there in 1 iteration.

$\Rightarrow$ ~~$2x - 2\delta < 2\delta$ $\Rightarrow$ take 1+ more iteration.~~

~~Because although B moves closer to O A will also move closer to O $\Rightarrow$ in cluster x all points are closest to A.~~

4b) $\Rightarrow$ $2x - 2\delta < 2\delta$ $\quad \delta << x$

$x - \delta < \delta$

$\Rightarrow$ $x < 2\delta$

if that is true we see point B will move closer to the origin but will not reach origin.

$\Rightarrow$ B be at $-\delta$ $\Rightarrow$

~~But~~ B is x away from closest right circle point. But A was previously 2δ away But now is $< 2\delta$ away $\Rightarrow$ A is closest to all points in right circle.

b) No, see a) explanation for when $x < 28$

5. a) Log likelihood that using parameters θ, how well test x will fit in the mixed gaussian $P(x;\theta)$

We're missing some data => easier to use EM. No, gives an estimate.

b) $P(x;\theta) = \frac{1}{2}\left(\frac{1}{6\sqrt{2\pi}} e^{\frac{-(x-6)^2}{2}} + \frac{1}{7\sqrt{2\pi}} e^{\frac{-(x-7)^2}{8}}\right)$

$P_1 = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(x-6)^2}{2}}$ $\quad P_2 = \frac{1}{7\sqrt{2\pi}} e^{-\frac{(x-7)^2}{8}}$

$\frac{P_2}{P_1} = \frac{\frac{1}{7} e^{-\frac{(x-6)^2}{2}}}{\frac{1}{6} e^{-\frac{(x-7)^2}{8}}} = \frac{1}{42}\left(e^{\left(-\frac{(x-6)^2}{2} + \frac{(x-7)^2}{8}\right)}\right)$

$= \frac{1}{42}\left(e^{\left(\frac{4(-x^2+12x-36)}{8} + \frac{x^2-14x+49}{8}\right)}\right)$

$1 < \frac{1}{42}\left(e^{\frac{-3x^2+34x-95}{8}}\right)$

$\ln(42) \cdot 8 < -3x^2 + 34x - 95$

$5.48 < x < 5.84 \sim 5$

$x^{(12)}$ or $x^{(13)}$

c) Forward.

a) θ_t^2 increase, θ_2^2 decreases.

P_t becomes more weighted as we move on.

c). θ_t will be larger in the end

& they should be switched?

b) $h_1 = \text{sign}(\theta' x)$

$= \left(\cos\left(\frac{\pi}{100}\right), \sin\left(\frac{\pi}{100}\right)\right)$

$h_1(x) = \text{sign}\left(\left(\cos\left(\frac{\pi}{100}\right), \sin\left(\frac{\pi}{100}\right)\right) \cdot x\right)$

$i = 1 \sim 50$, $151 \sim 199$ is some 1st & 3rd quad correct & some of 2nd & 4th.

$i = 51 \sim 100$, $101 \sim 150$ is 2nd & 4th & some of 1st & 3rd.

$\Rightarrow$ as you go up to the middle of each set the error decreases & increase.

Scanned by CamScanner