a.

$p(x) = x^2 - x - 1$

$p(0) = -1$ So, $p(0), p(1)$ classified $-ve$

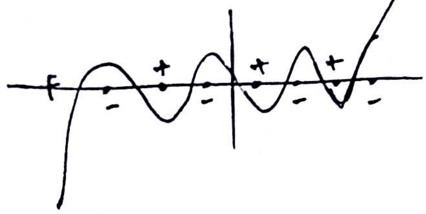
$p(1) = -1$

$p(2) = 1$ $p(2), P(3)$ classified $+ve$.

$p(3) = 5$

b.

We know that a d-degree polynomial will have d roots. ~~and d-1~~
So, the curve will cut the x-axis at d points. The most adverse configuration will be the alternate points configuration. and we can see we can correctly classify d+1 such points.



So, VC-dimension $\leq d+1$ —(i)

Again for VC-dimension $\geq d+1$,

Consider case for $d=1$,

we can correctly classify 2 points. ~~at most~~.

i.e. V-C dimension $= 2 \geq 1+1 \geq d+1$

Lets assume ~~Consider~~ a case for $d=k$ has V-C dimension $\geq k+1$.

From the reasoning discussed earlier, if we increase d to k+1,

it can classify one more point than d=k.

So, V-C dimension for $d = k+1$ is $(V\text{-}C_d + 1)$

$\geq k+1+1 \geq k+2$

So, Since we have V-C ≥ 2 for $k=1$,

By induction,

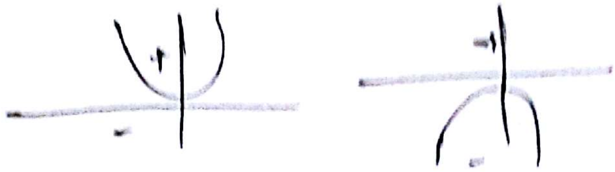
V-C $\geq k+1$ —(ii)

from (i) & (ii)

$\boxed{V\text{-}C = d+1}$

Scanned by CamScanner

c. If we restrict the coefficient to only positive, or say 1 we will decrease the v.c by 1 as we restrict the degree of freedom by 1. We can easily see in case of parabola.



Information Criteria.

(b). $BIC(M) = l - \frac{1}{2} p\log n$

$BIC(M_2) = l_2 - \frac{1}{2} \cdot 5p\log n$

$BIC(M_2) - BIC(M) > 0$

$l_2 - \frac{5}{2} p\log n - l_1 + \frac{1}{2} p\log n > 0$

$\Rightarrow \boxed{l_2 - l_1 \geq 2p\log n}$

(a) No, reducing 'n' to increase BIC score is not a good idea. Learning parameter with less number of data points doesn't ensure complete ~~team~~ ~~proper~~ learning as we may miss key characteristics.

(c). The penalty term is for number of parameters is larger in BIC than in AIC & hence more parameters the more penalty in BIC. So, simpler model will have less parameter and hence ~~the~~ less penalty. So, BIC prefers simpler models.

K-Means & K-Medoids

ⓐ 1st cluster. $\{(0,0),(0,6)\}$

center: $(0,-6)$

2nd cluster $\{\cancel{(0,-6)}\ (4,4),(-5,2)\}$ with center $(4,4)$

ⓑ

1st cluster. $\{(-5,2),(0,0),(4,4)\}$

center $\{(0,0)\}$

2nd cluster $\{(0,-6)\}$

center $(0,-6)$.

ⓒ. 1st center $\{(0,0)\ (0,-6)\}$ center $(0,-3)$

2nd cluster $\{(-5,2),(4,4)\}$

center $(-0.5,3)$

Algorithm

$$\sum_{i=4} \log \left[0.5\, N(\mu_1, \sigma_1^2) + 0.5\, N(\mu_2, \sigma_2^2) \right]$$

$$l(D;\theta) = \sum_i \sum_y P(y|i) \log \left[\frac{P(x^{(i)}, y; \theta)}{P(y|i)} \right]$$

$$= \sum_i \sum_y P(y|i) \log \left[\frac{\cancel{P(x^{(i)}}\, \pi\, N(x^{(i)}, \theta)}{P(y|\theta)} \right]$$

$$= \sum_i \sum_y P_y \cdot \frac{N(x; \mu^{(y)}, \theta_y^2 I)}{P(x|\theta)} \cdot \log \left[\frac{\frac{\cancel{\pi\, N(x^{(i)}, \theta}}{\cancel{\pi_y N(x, \mu^{(y)}}}}{\sum_{i=1}^{k} \pi_y N(x}\right.$$

$$\hat{l}(D;\theta) = \sum_i \sum_y 1 \cdot \log \sum_{i=1}^{k} \pi\, N(x; \mu^y, \theta_y^2 I)$$

Scanned by CamScanner

$$\frac{P(y=2 \mid x^{(i)}, \theta_0)}{P(y=1 \mid x^{(i)}, \theta_0)} > 1$$

$$\Leftrightarrow \frac{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}(x-\mu_2)^2 / \sigma_2^2\right)}{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}(x-\mu_1)^2 / \sigma_1^2\right)} > 1$$

$$\Leftrightarrow \frac{\frac{1}{2}}{\frac{1}{1}} \cdot \frac{\exp\left(-\frac{1}{2}(x-7)^2 / 4\right)}{\exp\left(-\frac{1}{2}(x-6)^2 / 1\right)} > 1$$

$$\Leftrightarrow \frac{1}{2} \exp\left(-\frac{1}{2} \cdot \frac{(x-7)^2 - 4(x-6)^2}{4}\right) > 1$$

$$\Leftrightarrow \frac{1}{2} \exp\left[\frac{1}{8} \cdot (-x^2 + 14x + 49 + 4x^2 - 48x + 36 \times 4)\right] > 1$$

$$\Leftrightarrow \frac{1}{2} \exp\left[\frac{1}{8}\left(3x^2 - 34x + 36 \times 4 + 49\right)\right] > 1$$

$$\Rightarrow \log\frac{1}{2} + \frac{1}{8}(x-5)(3x-19) > 0$$

$$\Rightarrow \boxed{x = 4.15, \; x = 7.18} \qquad x \in [4.15, 7.18]$$

So, the point x^3, x^4 are more likely to assign to class y_1. ~~then~~ x_0, x_1, x_2, x_3 will have $P(y=2 \mid x^{(i)}, \theta_0) > P(y=$

d.

Both ~~direct~~ Gaussians will move to left. Since the mean of both clusters is greater than actual values.

e. σ_1 will increase & σ_2 will decrease.

f. Cluster 2 will have larger variance since it has 4 values ~~and~~ ranging from -1 to 4. But Cluster 1 has only 2 points, ~~with~~ with range 5