ML QUIZ ① Kv2154

Problem 1

$$f(x_i, \theta_0, \theta_1, \theta_2 \ldots \theta_p) = \sum_{p=1}^{P} \theta_p x_i^p + \theta_0$$

Empirical risk $R(\theta)$

$$R(\theta) = \frac{1}{2N}\|y - X\theta\|^2 + \theta^T\theta + \alpha^T\theta$$

$$\frac{\partial R(\theta)}{\partial(\theta)} = \frac{\partial}{\partial\theta}\left[\frac{1}{2N}\|y-X\theta\|^2 + \theta^T\theta + \alpha^T\theta\right]$$

$$= \frac{\partial}{\partial\theta}\left[\frac{1}{2N}(y-X\theta)^T(y-X\theta) + \theta X^T X\theta + \theta^T\theta + \alpha\theta\right]$$

$$= \frac{\partial}{\partial\theta}\left[\frac{1}{2N}(y^Ty - 2y^TX\theta + \theta^Tx^Tx\theta)\right] + \frac{\partial}{\partial\theta}(\theta^T\theta)$$

$$+ \frac{\partial}{\partial\theta}(a^T\theta)$$

$$= \frac{1}{2N}(0 - 2y^Tx + 2\theta^Tx^Tx) + 2\theta^T + a$$

Finding Optimal Solution

$$\frac{1}{2N}(-2y^Tx + 2\theta^Tx^Tx) + 2\theta^T + \alpha = 0$$

KV2154

$$-\frac{1}{N} Y^T X + \frac{1}{N} \Theta^T X^T X + 2\Theta^T + \alpha^T = 0$$

$$\frac{1}{N} \Theta^T [X^T X + 2NI] = \frac{1}{N} [Y^T X - N\alpha^T]$$

$$\Theta^T = (Y^T X - N\alpha^T)(X^T X + 2NI)^{-1}$$

$$\Theta^* = ((Y^T X - N\alpha^T)^T ((X^T X + 2NI)^{-1})^T$$

$$\Theta^* = (X^T Y - N\alpha)(X^T X + 2NI)^{-1}$$

Problem 2

$w = xy^2$ $\qquad x = r\cos\theta,\ y = r\sin\theta$

$r = 2,\ \theta = \pi/4$

$\cancel{r^3 \cos\theta \sin^2\theta}$

$$\frac{\partial w}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial r}$$

$$= \frac{\partial}{\partial x} xy^2 \cdot \frac{\partial}{\partial r} r\cos\theta + \frac{\partial}{\partial y} xy^2 \cdot \frac{\partial}{\partial r} r\sin\theta$$

$$= y^2 \cdot \cos\theta + 2xy \cdot \sin\theta$$

$$= r^2 \sin^2\theta \cos\theta + 2r^2 \sin^2\theta \cos\theta$$

$$= 3r^2 \sin^2\theta \cos\theta$$

$$= 3 \cdot 4 \cdot \sin^2\left(\frac{\pi}{4}\right) \cdot \cos\left(\frac{\pi}{4}\right)$$

$$= 4.24264$$

Problem (3)

$$f(x|\theta) = \frac{1}{\theta} x^{(1-\theta)/\theta}$$

The maximum likelihood is given by

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} x^{\left(\frac{1-\theta}{\theta}\right)}$$

The log likelihood is given by

$$l(\theta) = \log \left[\prod_{i=1}^{n} \frac{1}{\theta} x_i^{\left(\frac{1-\theta}{\theta}\right)} \right]$$

$$= \sum_{i=1}^{n} \log \left(\frac{1}{\theta^n} x_i^{\left(\frac{1-\theta}{\theta}\right)} \right)$$

$$= \log \left(\frac{1}{\theta^n} \right) + \left(\frac{1-\theta}{\theta} \right) \sum_{i=1}^{n} \log x_i$$

$$l(\theta) = \log \left(\frac{1}{\theta^n} \right) + \frac{1}{\theta} \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \log x_i$$

for MLE

$$\frac{\partial l(\theta)}{\partial(\theta)} = 0$$

KV2154

$$\frac{\partial}{\partial \theta} -n\log\theta + \frac{1}{\theta}\sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \log x_i = 0$$

$$-\frac{n}{\theta} - \frac{1}{\theta^2}\sum_{i=1}^{n} \log x_i = 0$$

$$\frac{1}{\theta^2}\sum_{i=1}^{n} \log x_i = -\frac{n}{\theta}$$

$$\hat{\theta}_{MLE} = -\frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$x_1 = 0.10,\ x_2 = 0.22,\ x_3 = 0.55,\ x_4 = 0.36$$

$$= -\frac{1}{4}\{\log(0.10) + \log(0.22) + \log(0.54) + \log(0.36)\}$$

$$= -\frac{1}{4}(-2.3026 - 1.5141 - \overset{0.6162}{\cancel{0.6162}} - 1.0217)$$

$$= -\frac{1}{4}(-5.4546)$$

$$= 1.36365$$

kv2154

Problem 4

$f(x,y) = xe^y$, Constraint $x^2+y^2=2$

given , $f(x,y) = xe^y$
$g(x,y) = x^2+y^2=2$
$\nabla f = \lambda \nabla g$
$\langle e^y, xe^y\rangle = \langle 2\lambda x, 2\lambda y\rangle$, So we get Eqn

$e^y = 2\lambda x$ $\qquad$ $xe^y = 2\lambda y$
$\lambda = \frac{e^y}{2x}$ $\qquad$ $\lambda = \frac{xe^y}{2y}$

$\frac{e^y}{2x} = \frac{xe^y}{2x}$

$y = x^2$ —— (1)

Sub (1) in $g(x,y)$

$y+y^2=2$
$y^2+y-2=0$
$y^2+2y-y-2=0$
$y(y+2)-1(y+2)=0$
$(y+2)(y-1)=0$
$y=1,-2$ $\qquad$ $x = 1, \sqrt{2}i$ (Not Valid)

$\therefore \; y = 1$

$x = 1, -1$ (from 1)

$f(1,1) = (1)\,e^{(1)} = e = 2.718$

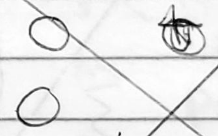
$f(-1,1) = (-1)\,e^{1} = -e = -2.718$

extreme values are $2.718, -2.718$

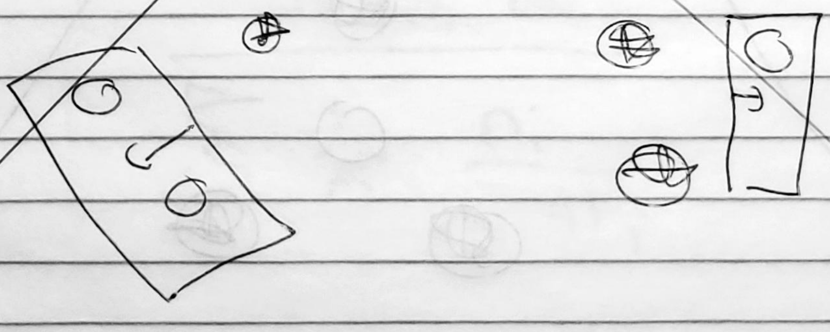
Problem 8

The VC Dimension of an axis-aligned rectangles in a 2D plane is 3.

→ Consider a space in which there are 3 ~~diff~~ data points



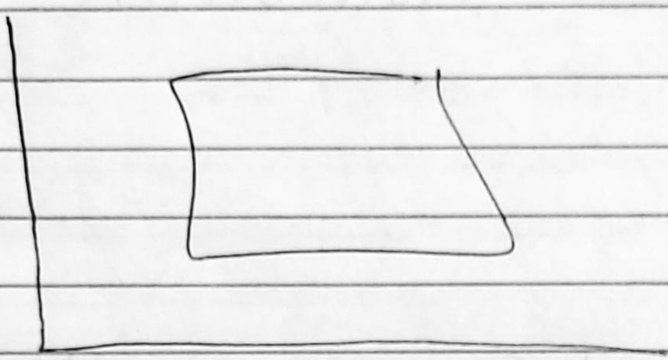
Now, for a given function $f(x, \theta)$, you can correctly classify the following distribution



KVZ154

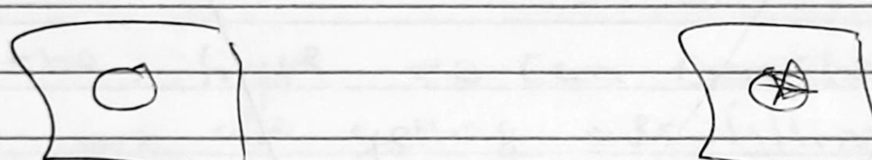
Problem 8

→ The axis align rectang will be

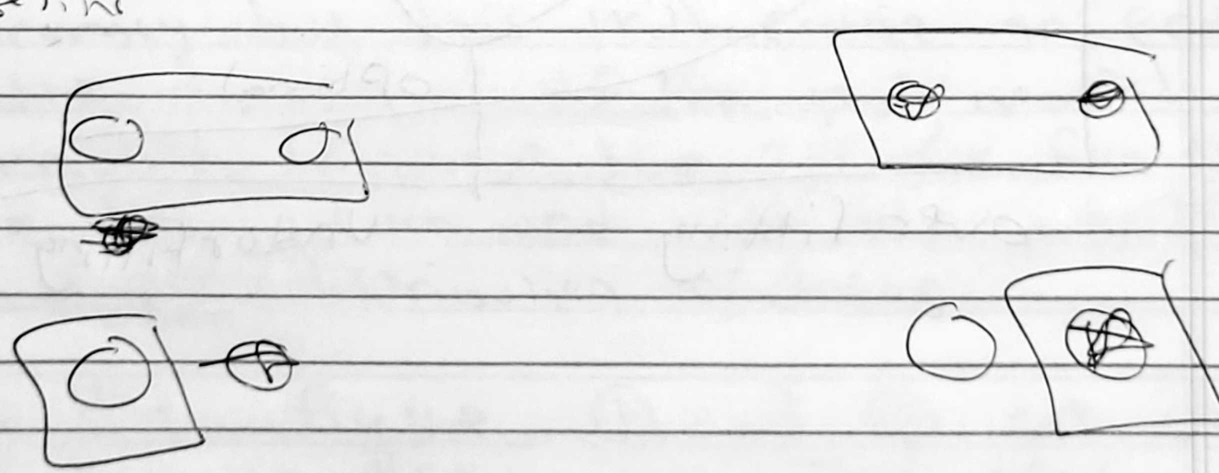


Consider 2 data points O and ⊗

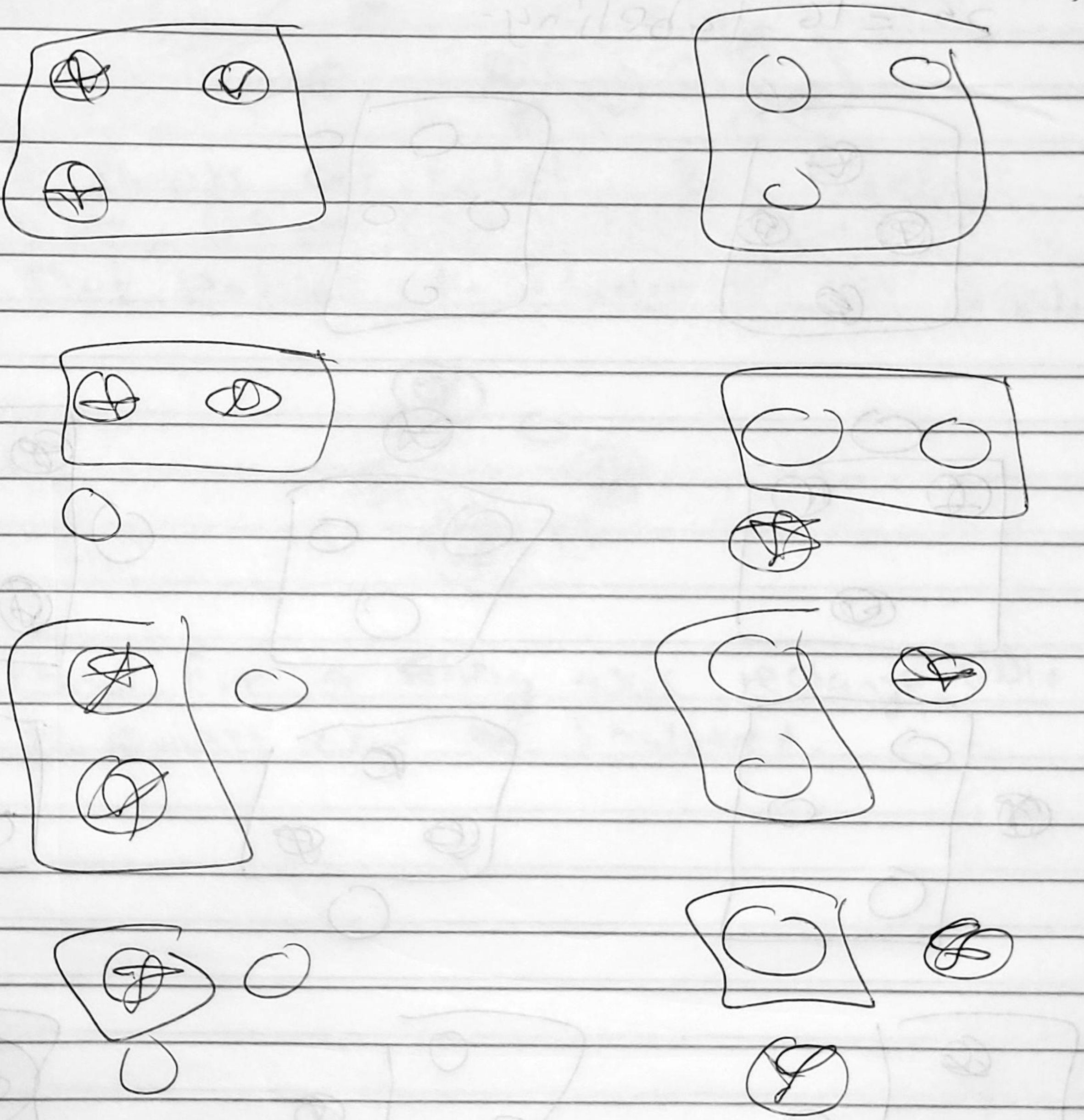
So for 1 point there are upto $2^1 = 2$ labelling.



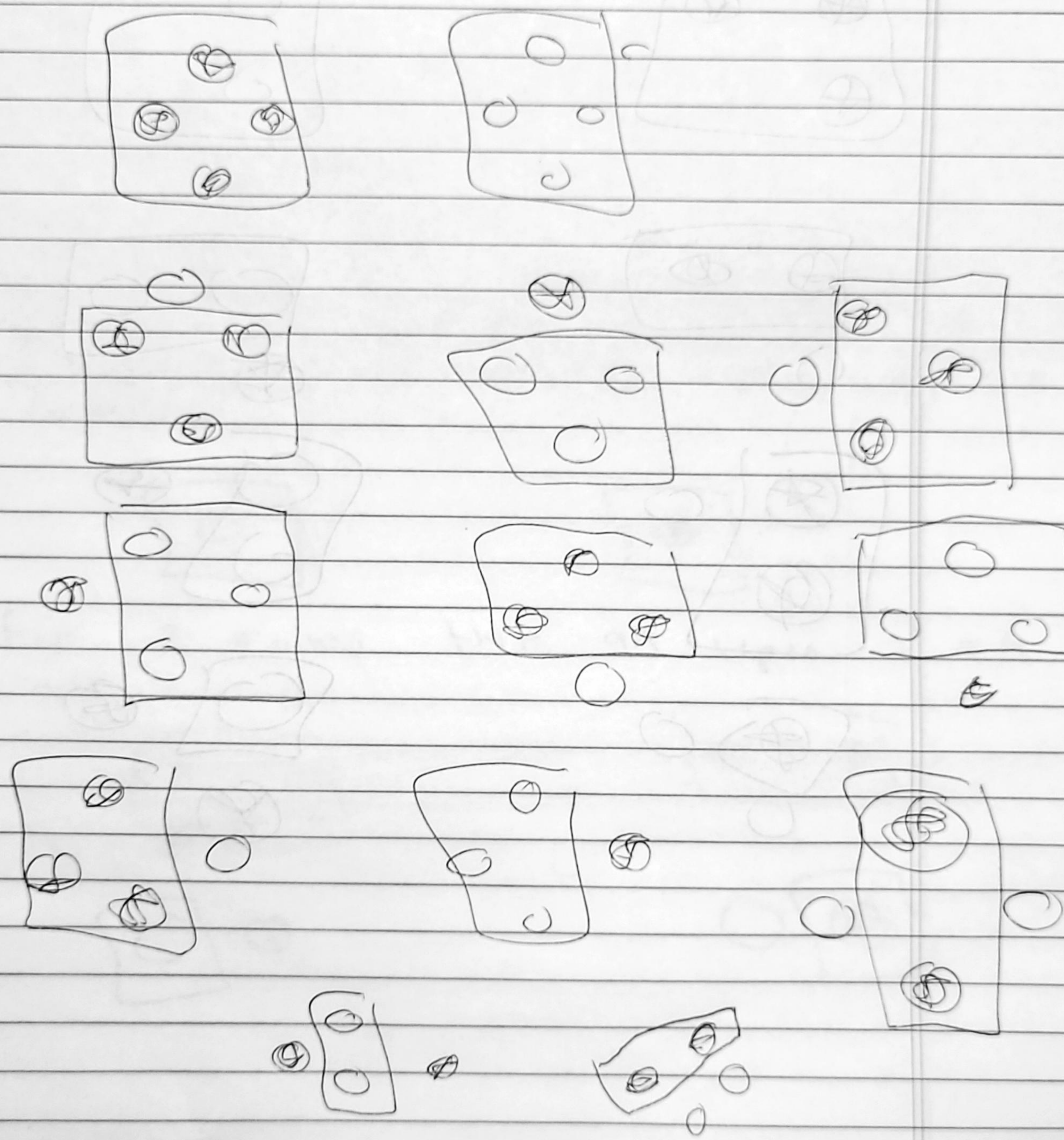
for 2 points there are upto $2^2 = 4$ labelling



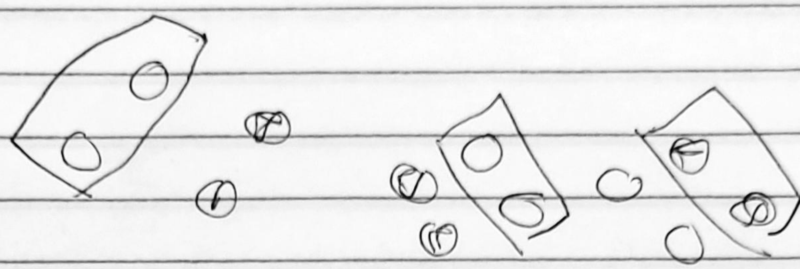
for 3 points upto $2^3 = 8$ labeling



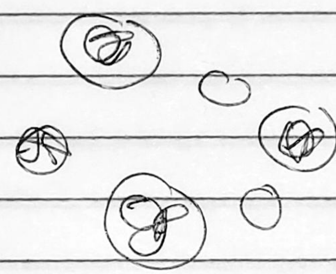
for 4 points numbers of possibilities
$2^4 = 16$ labeling.



KV2154 1(b)



5 points are total 32 labelings
Now, there are cases where it is not possible to shatter all the points



Thus for a saling axis rectange upto 4 points can be shattered.

KV2IS4

~~Thus, the VC dimensions of axis-aligned rectangles is atmost 3 as not all sets of 4 points can be fully shattered.~~

Problem 6

boxes $= m$ Probability of selecting $i = \frac{1}{m}$

ith box $= n_i$ apples r_i oranges

Probability of selecting apple in ith box

$P(A/B_i) = \frac{n_i}{n_i + r_i}$

$$P(B_i/A) = \frac{P(A/B_i) \cdot P(B_i)}{\sum_{i=1}^{m} P(B_i) \cdot P(A/B_i)}$$

$$= \frac{\frac{n_i}{n_i + r_i} \cdot \frac{1}{m}}{\sum_{i=1}^{m} \frac{1}{m} \cdot \frac{n_i}{n_i + r_i}}$$

Probability that apple is from box j

$$P(B_j/A) = \frac{\frac{n_j}{n_j+r_j} \cdot \frac{1}{m}}{\sum_{i=1}^{m} \frac{n_i}{n_i+r_i} \cdot \frac{1}{m}}$$

$$P(B_j/A) = \frac{\frac{n_j}{n_j+r_j}}{\frac{n_i}{n_i+r_i}}$$

Problem 7

$$M(f(x;w)) = \sum_{j=0}^{M} w_j x^j$$

where $w = [w_0 \ w_1 \dots w_M]^T$

Part 1,

from the figure, the most reasonable choice for M is M = 3, as it properly fits the distribution of data without any stray behavior of the function; which might add to the loss down the line.

From the figure we can conclude that at M=9, we are getting over fitting. As the said function transcends every point in the distribution and fits it with higher accuracy but this will cause an error as the dimension of the polynomial distribution becomes too specific for the given distribution and will add up to the loss of other distribution of datas.

From the figure, 1 and 2 are under fitted for the distribution. The reason being that the curve of the graph 1 and 2

KV2154

12

Doesnt ~~satisfend~~ fit most of the data points thus ~~creating~~ generating higher loss of the distribution.

Part B

Squared loss

$$L(y_i, f(x_i; w)) = \frac{1}{2}(y_i - f(x_i; w))^2$$