

Problem 1

Question 1.

60 healthy samples 40 unhealthy samples
 FA: 40H 20UH 20H 20UH
 FB: 30H 10UH 30H 30UH

$$\begin{aligned}\Delta H(F_A) &= H(6/10, 4/10) - 6/10 \log_2(4/6, 2/6) - 4/10 \log_2(2/4, 2/4) \\ &= -\frac{6}{10} \log_2\left(\frac{6}{10}\right) - \frac{4}{10} \log_2\left(\frac{4}{10}\right) - \frac{6}{10} \left(-\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right)\right) \\ &\quad - \frac{4}{10} \left(-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right)\right) \\ &= 0.019973\end{aligned}$$

$$\begin{aligned}\Delta H(F_B) &= H(6/10, 4/10) - \frac{6}{10} \log_2(3/4, 1/4) - \frac{4}{10} \log_2(3/6, 3/6) \\ &= -\frac{6}{10} \log_2\left(\frac{6}{10}\right) - \frac{4}{10} \log_2\left(\frac{4}{10}\right) - \frac{6}{10} \left(-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right)\right) \\ &\quad - \frac{4}{10} \left(-\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right)\right) \\ &= 0.0465\end{aligned}$$

as F_B has more certainty it is a better classifier

Problem 2

Question 2

0.7, 0.5, 0.9

0.2, 0.3, 0.4

+1 -1 +1

$$a) E = \frac{0.6}{0.9} \quad \alpha = 0.5 \ln \left(\frac{1 - \frac{0.6}{0.9}}{\frac{0.6}{0.9}} \right) = -0.34657$$

$w_1: w_1 = 0.2$

$$\alpha = -0.34657$$

$$w_1 = 0.2e$$

$$y = (-1)$$

$$= 0.1414218$$

$$F(x) = (1)$$

$w_2: w_2 = 0.3$

$$\alpha = -0.34657$$

$$w_2 = 0.3e$$

$$y = -1$$

$$= 0.4242625$$

$$F(x) = -1$$

$w_3: w_3 = 0.4$

$$w_3 = 0.4e$$

$$\alpha = -0.34657$$

$$= 0.282843$$

$$y = -1$$

$$F(x) = +1$$

Problem 3

b) The coefficient of the classifier is \times previously calculated -0.34657 .

Question 3

Length: 12 Weight: 122g Colour: Red

$$P(O|D) = \frac{P(D|O)P(O)}{P(D)}$$

$$= 0$$

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{\left(\frac{20}{50}\right)\left(\frac{30}{50}\right)\left(\frac{25}{50}\right)\left(\frac{50}{275}\right)}{\frac{6}{275}} = \frac{6}{275} = 0.0218$$

$$P(M|D) = \frac{\left(\frac{45}{75}\right)\left(\frac{55}{75}\right)\left(\frac{25}{75}\right)\left(\frac{75}{275}\right)}{\frac{1}{25}} = \frac{1}{25} = 0.04$$

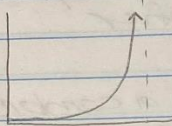
$$P(P|D) = \frac{\left(\frac{40}{80}\right)\left(\frac{35}{80}\right)(1)\left(\frac{80}{275}\right)}{\frac{7}{110}} = \frac{7}{110} = 0.063$$

\therefore The classifier should output pomegranite!

Problem 4

Question 4

a) The upper bound would be infinity. If two points were taken infinitely far from each other in a $k=1$ space the error will be infinite. This can be extended without loss of generality to any number of points infinitely far from a given centroid.



is what the error would look like

The lower bound is 0, this occurs when the centroid is located at the centre of mass of all datapoints.

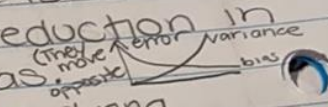
Problem 5

Question 5

a) True. Using a lower k -value than appropriate will result in the model overfitting to satisfy the conditions of the k -nearest neighbours.

b) False. We want a higher number of true positives. An equal amount would give a 50% accuracy which is the same as a random classifier.

c) True. It must be linearly separable by a combination of classifiers to obtain a training error of 0 we can achieve this using enough classifiers.

d) False. It does not increase bias & variance for a non-linear regressor. We have a reduction in variance but an increase in bias. 

e) False: we can use layers to extract features during, however these are not used as inputs for training the raw image is the input. (without loss of generality)

f) False. The theorem states that the averaging of performance over all problems will perform the same as another model.