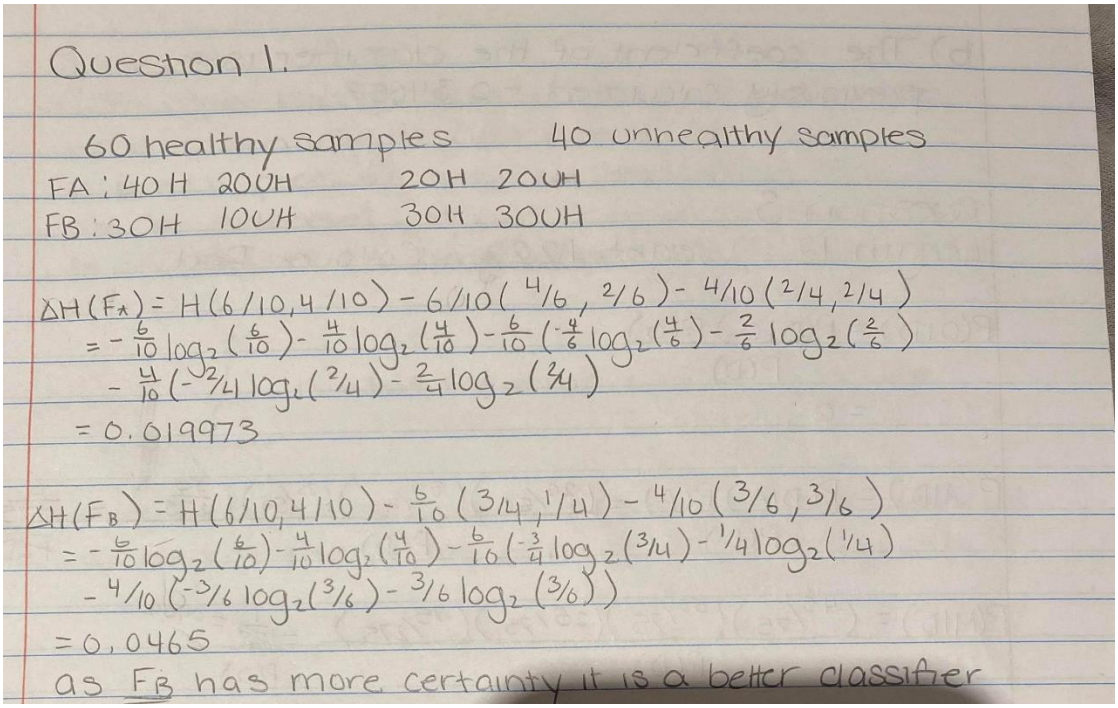
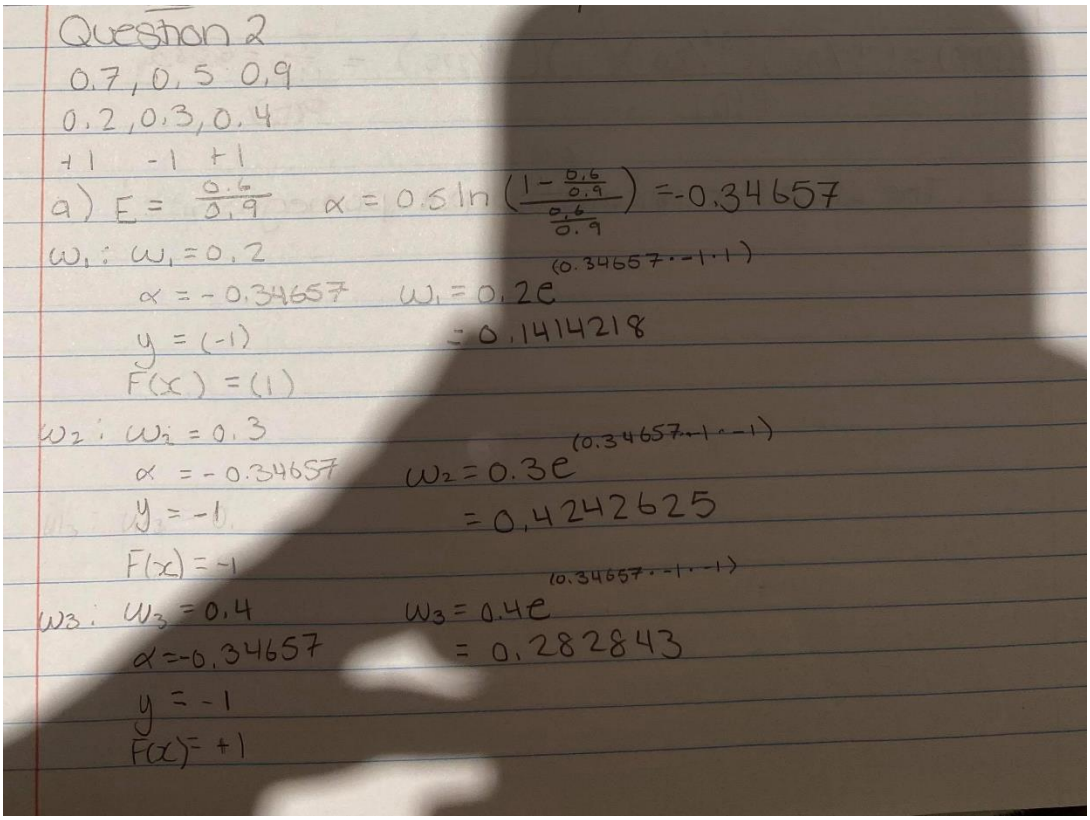
Problem 1



Problem 2



Problem 3

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

Question 3

Length: 12 Weight: 122g Colour: Red

$$P(O|D) = \frac{P(D|O)P(O)}{P(D)} = 0$$

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

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

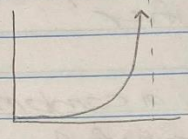
$$P(P|D) = \frac{(40/80)(35/80)(1)(80/275)}{P(D)} = \frac{7}{110} = 0.063$$

∴ The classifier should output pomegranite!

Problem 4

Question 4

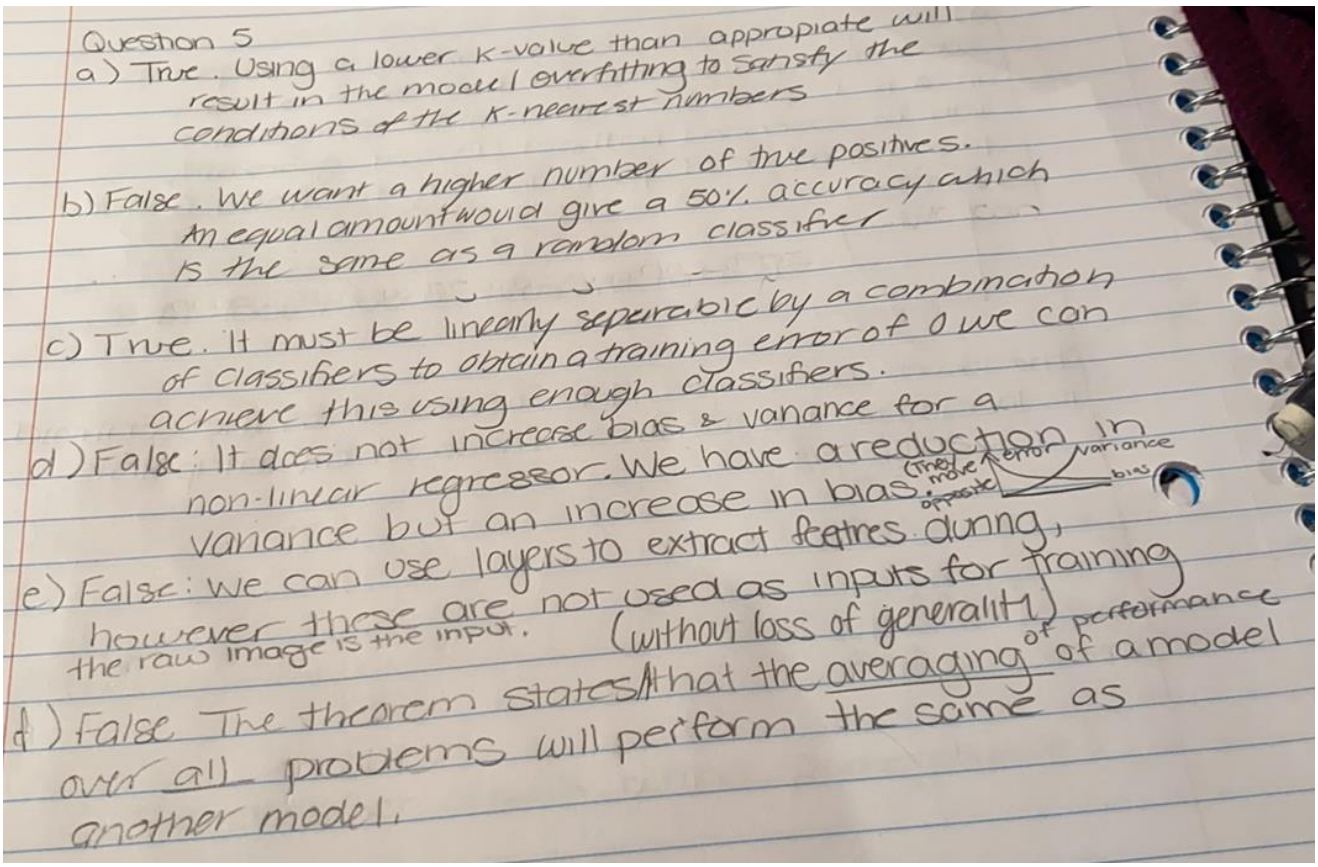
a) The upper bound would be infinity. If two points were taken infinately far from each other in a k=1 space the error will be infinite. This can be extended without loss of generality to any number of points infinety far from a given centroid.



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 appropiate will result in the model overfitting to satisfy the conditions of the k-nearest numbers
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.
(The more complex)
error
variance
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.
f) False The theorem states that the averaging (without loss of generality) of performance over all problems will perform the same as another model.