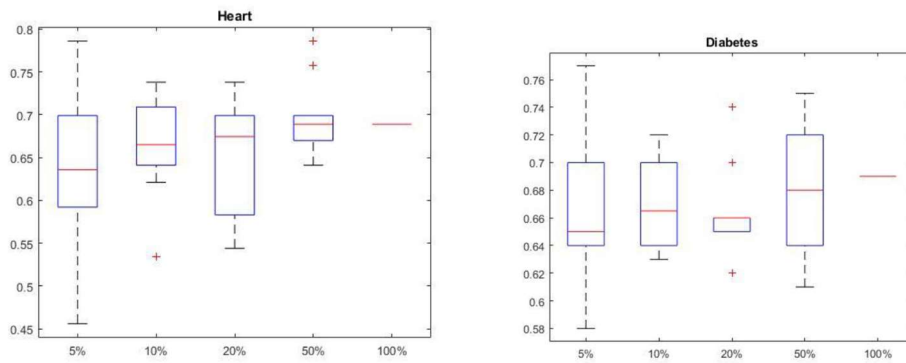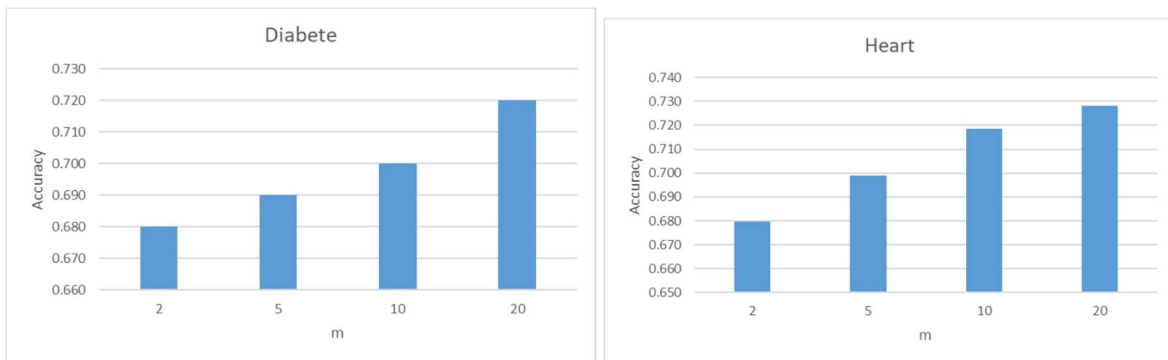Part 2

Part 3

Part 4

1. Information Theory

(a) If $H(X|Y) = H(X)$ then X and Y are independent.
   True:

$$I(X;Y) = H(X) - H(X|Y)$$

If $I(X;Y) = 0$ then $H(X) = H(X|Y)$. We can write:

$$I(X;Y) = D(\ P_{X,Y} \ ||\ P_X P_Y\ ) = 0$$

$D(Q||P) = 0$ iff $P(x) = Q(x)\ \forall x$, therefore $P_{X,Y}(x,y) = P_X(x)P_Y(y)$
for every x, y and as result $X \perp Y$.

2. Standardize the data set with four points in 2 dimension: (7, 7),(3, 7),(3, 3),(7, 3);
   $u_1 = 5$, $u_2 = 5$, $\sigma_1 = 2$, $\sigma_2 = 2$,
   (7, 7)->(1, 1)
   (3, 7)->(-1,1)

(3, 3)->(-1, -1)
(7, 3)->(1, -1)

3. KNN
   (1) 1NN: x3 is the closest, then its label is 1.
   (2) 2NN: x2, x3, x4 are the closest, then its label is 1.
   (3) 3NN: We need to get the reciprocal value of the distance, then its label is 1.

4. Performance Measurements
   (1) Accuracy = (TP+TN)/(TP+FP+FN+TN) = (76+82)/(76+18+24+82) = 0.79
   (2) Error = 1-Accuracy = 0.21
   (3) True positive rate = TP/actual pos = TP/(TP+FN) = 76/(76+24) = 0.76
   (4) False positive rate = FP/actual neg = FP/(TN+FP) = 18/(18+82) = 0.18
   (5) Precision = TP/predicted pos = TP/(TP+FP) = 76/(76+18) = 0.8085
   (6) Recall = TP/actual pos = TP/(TP+FN) = = 76/(76+24) = 0.76