Homework 2:
Written Exercise Part

>>ARPIT JAIN<<
>>ajain74<<

1 Information Theory [25/4 pts]

Suppose X, Y are two random variables taking values in a discrete finite set V. Let $H(Y)$ denote the entropy of Y, and let $H(Y|X)$ denote the conditional entropy of Y conditioned on X. Prove that if X, Y are independent, then $H(Y) = H(Y|X)$.

As we know, $H(X) = -\sum_{x\in\{x_i,x_2,..x_k\}} P(X=x)\log_2 P(X=x)$
and, $H(Y|X) = -\sum_{x\in\{x_i,x_2,..x_k\}} P(X=x)H(Y|X=x)$
now,

$$H(Y|X=x) = -\sum_{y\in\{y_i,y_2,..y_k\}} P(Y=y|X=x)\log_2 P(Y=y|X=x)$$

given, X, Y are independent,

$$P(Y=y|X=x) = P(Y=y)$$

$\implies H(Y|X=x) = -\sum_{y\in\{y_i,y_2,..y_k\}} P(Y=y)\log_2 P(Y=y)$
$\implies H(Y|X=x) = H(Y)$

$\implies H(Y|X) = \sum_{x\in\{x_i,x_2,..x_k\}} P(X=x)H(Y)$
as, $\sum_{x\in\{x_i,x_2,..x_k\}} P(X=x) = 1$

$$H(Y|X) = H(Y)$$

2 Standardizing Numeric Features [25/4 pts]

Standardize the data set with four points in 2 dimension: $(7,7), (3,7), (3,3), (7,3)$.

Given, $(7,7), (3,7), (3,3), (7,3)$
Let us first calculate the mean and standard deviation for this set.

Mean(x) = $\mu(x) = \frac{7+3+3+7}{4}$
$\implies = \frac{20}{4} = 5$

Mean(y) = $\mu(y) = \frac{7+7+3+3}{4}$
$\implies = \frac{20}{4} = 5$

Std. Deviation (x) = $\sigma(x) = \sqrt{\frac{(7-5)^2+(3-5)^2+(3-5)^2+(7-5)^2}{4}}$
$\implies = \sqrt{\frac{4+4+4+4}{4}} = 2$

Std. Deviation (y) = $\sigma(y) = \sqrt{\frac{(7-5)^2+(3-5)^2+(3-5)^2+(7-5)^2}{4}}$
$\implies = \sqrt{\frac{4+4+4+4}{4}} = 2$

For standardization of a set of points: $(x')_i = (x_i - \mu_i)/\sigma$

$\implies x_1 = \frac{7-5}{2} = 1, y_1 = \frac{7-5}{2} = 1$
$\implies x_2 = \frac{3-5}{2} = 1, y_1 = \frac{7-5}{2} = 1$
$\implies x_3 = \frac{3-5}{2} = 1, y_1 = \frac{3-5}{2} = -1$
$\implies x_4 = \frac{7-5}{2} = 1, y_1 = \frac{3-5}{2} = -1$

Standardized points:

$$(1,1), (-1,1), (-1,-1), (1,-1)$$

3 k-Nearest Neighbors [25/4 pts]

Consider the training data set $x_1 = (7,7), y_1 = 0; x_2 = (3,7), y_2 = 1; x_3 = (3,3), y_3 = 1; x_4 = (7,3), y_4 = 2$. Suppose the Manhattan distance is used. What is the label for $x = (0,0)$ in the following settings? Show the calculation steps.

1. 1-nearest neighbors.
2. 3-nearest neighbors.
3. 3-nearest neighbors, distance weighted. The weight for the i-th neighbor z is $1/d(x,z)^2$.

Manhattan Distance = $|x - x_i|$ for x_1, x_2, x_3, x_4
$d(x,x_1) = |0-7| + |0-7| = 14$
$d(x,x_2) = |0-3| + |0-7| = 10$
$d(x,x_3) = |0-3| + |0-3| = 6$
$d(x,x_4) = |0-7| + |0-3| = 10$

1. 1-nearest neighbors. For this, d_3 is the lowest distance of all.
   $\implies$ So nearest neighbors $= x_3$.
   $\implies$ Label for $x = y_3 = 1$
2. 3-nearest neighbors. For this, d_2, d_3, d_4 are the lowest distance of all.
   $\implies$ So 3nearest neighbors x_2, x_3, x_4.
   $\implies$ Label for $x = \sum_{i=2}^{4} \gamma(v, y_i); v \in \{1,2,3\}$
   $\implies$ Label for $x = 1$
3. 3-nearest neighbors, distance weighted.

   As we know, $w_i = \frac{1}{d(x,x_i)^2}$
   $w_2 = \frac{1}{d(x,x_2)^2} = \frac{1}{10^2} = \frac{1}{100}$
   $w_3 = \frac{1}{d(x,x_3)^2} = \frac{1}{6^2} = \frac{1}{36}$
   $w_4 = \frac{1}{d(x,x_4)^2} = \frac{1}{10^2} = \frac{1}{100}$

   $\implies$ $l$label = $\begin{bmatrix} \frac{1}{100} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{36} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{100} \\ 0 \end{bmatrix}$
   = 1

4 Performance Measurements [25/4 pts]

Consider the following confusion matrix for 2 classes.

	actual positive	actual negative
predict positive	76	18
predict negative	24	82

Compute the accuracy, error, true positive rate, false positive rate, precision, and recall.

given,
True Positive(TP) = 76
False Positive(FP) = 18
True Negative(TN) = 82
False Negative(FN) = 24

Accuracy = $\frac{TP+TN}{TP+FP+FN+TN} = \frac{76+82}{200} = 0.79$

Error = 1 - Accuracy $= 1 - 0.79 = 0.21$

True positive rate = $\frac{TP}{actual\,positive} = \frac{TP}{TP+FN} = \frac{76}{76+24} = 0.76$

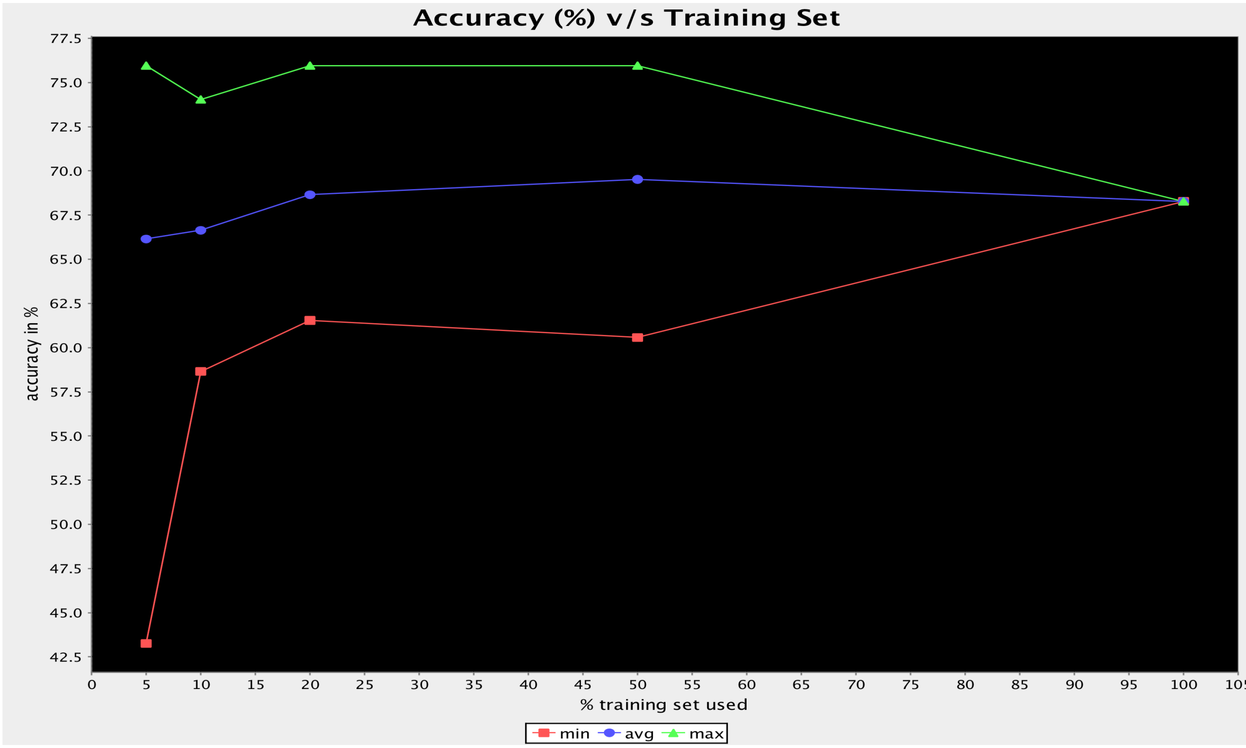
False positive rate = $\frac{FP}{actual\,negative} = \frac{FP}{FP+TN} = \frac{18}{18+82} = 0.18$

Precision = $\frac{TP}{predicted\,negative} = \frac{TP}{TP+FP} = \frac{76}{76+18} = 0.80$

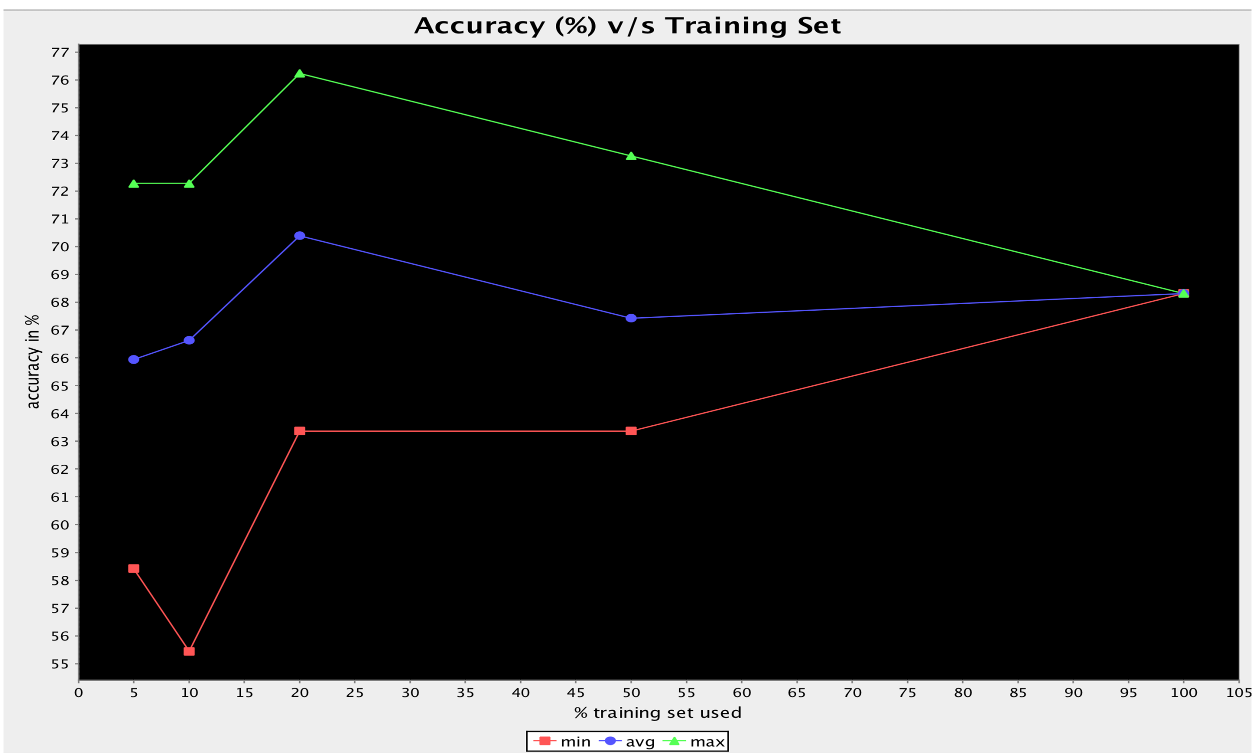
Recall = $\frac{TP}{actual\,positive} = \frac{TP}{TP+FN} = \frac{76}{76+24} = 0.76$

5 Graphs for Part 2 and 3

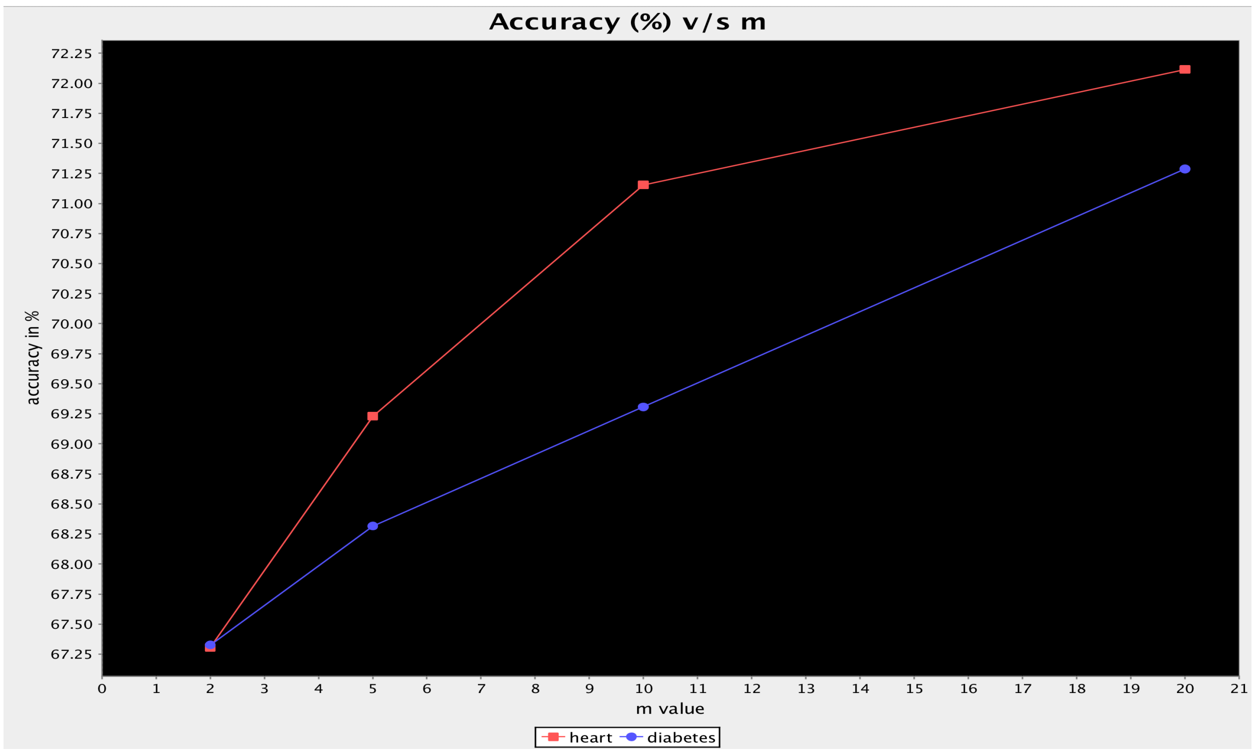
In both the figures (on next page) for part 2, the accuracy converges as the % of training data used is 100. The same is because, with 100% data, the model is always constant and hence will provide the same set of results for the given testing data.



for HEART data



for DIABETES data


Accuracy (%) v/s m
accuracy in %
72.25
72.00
71.75
71.50
71.25
71.00
70.75
70.50
70.25
70.00
69.75
69.50
69.25
69.00
68.75
68.50
68.25
68.00
67.75
67.50
67.25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
m value
heart
diabetes