

Name - Ankush Vijay Israney
Student ID - 14057308
CS 613 - Assignment 3 report

PART-1 [Theory Part]

1:

A - Sample Entropy

$$H(Y) = -\frac{4}{7} * \log_2 \frac{4}{7} - \frac{3}{7} * \log_2 \frac{3}{7} = 0.9851 \quad (1)$$

B - Information Gains

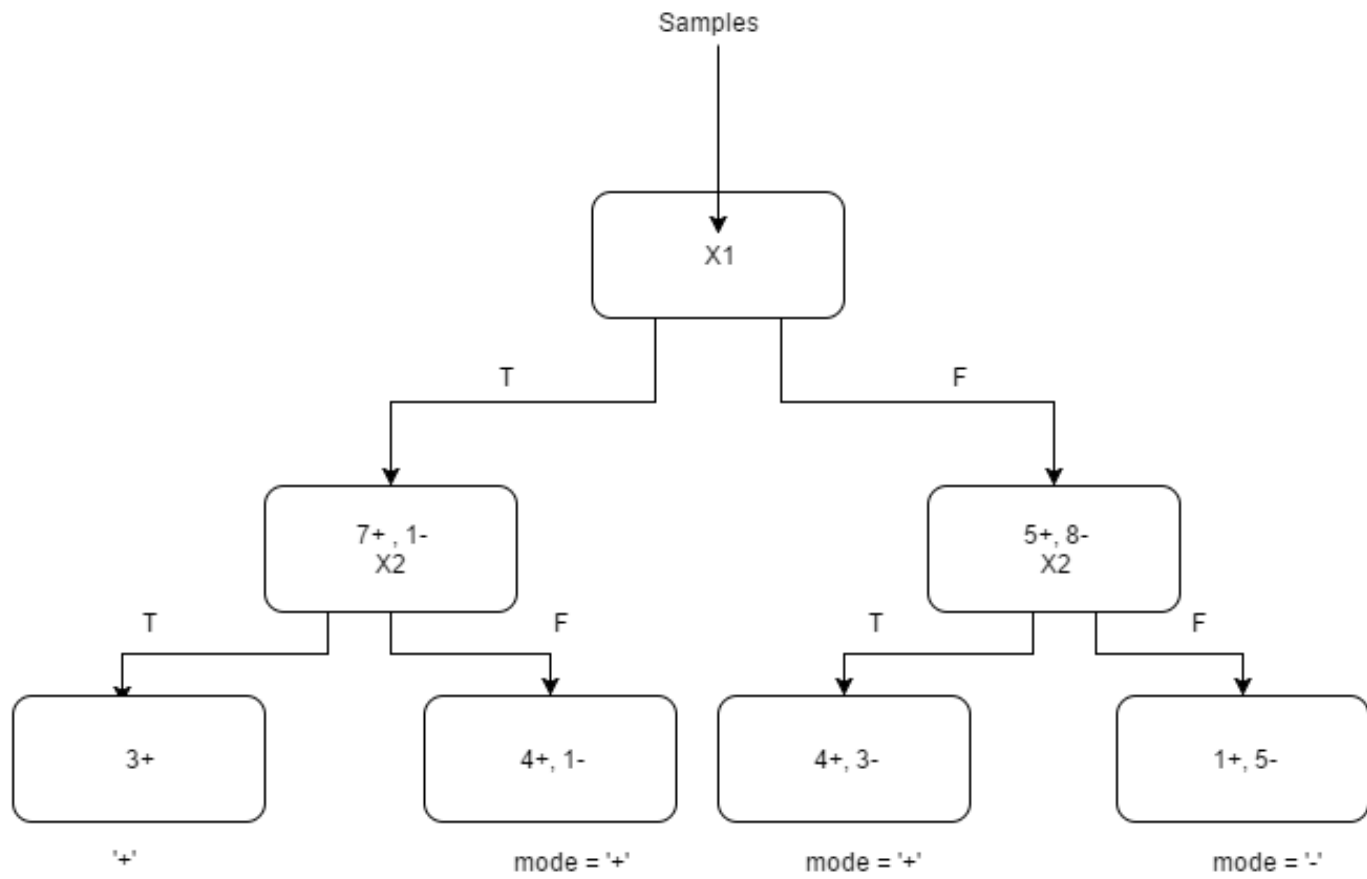
$$H(x_1) = \frac{8}{21}(-\frac{7}{8} * \log_2 \frac{7}{8} - \frac{1}{8} * \log_2 \frac{1}{8}) + \frac{13}{21}(-\frac{5}{13} * \log_2 \frac{5}{13} - \frac{8}{13} * \log_2 \frac{8}{13}) = 0.802 \quad (2)$$

$$Information - Gain(x_1) = H(Y) - H(x_1) = 0.1831 \quad (3)$$

$$H(x_2) = \frac{10}{21}(-\frac{7}{10} * \log_2 \frac{7}{10} - \frac{3}{10} * \log_2 \frac{3}{10}) + \frac{11}{21}(-\frac{5}{11} * \log_2 \frac{5}{11} - \frac{6}{11} * \log_2 \frac{6}{11}) = 0.9404 \quad (4)$$

$$Information - Gain(x_2) = H(Y) - H(x_2) = 0.0447 \quad (5)$$

C - Decision Tree



2:

A - Class Priors

$$P(A = Yes) = 0.6, P(A = No) = 0.4 \quad (6)$$

B - Gaussian Parameters

X =

216	5.68	yes
69	4.78	yes
302	2.31	no
60	3.16	yes
393	4.2	no

X-standardized =

$$\begin{bmatrix} 0.0551 & 1.2477 & yes \\ -0.9572 & 0.5688 & yes \\ 0.6473 & -1.2945 & no \\ -1.0192 & -0.6533 & yes \\ 1.2740 & 0.1313 & no \end{bmatrix}$$

$$\mu_{yes1} = -0.6404, \sigma_{yes1} = 0.6031 \quad (7)$$

$$\mu_{yes2} = 0.3877, \sigma_{yes2} = 0.9633 \quad (8)$$

$$\mu_{no1} = 0.9606, \sigma_{no1} = 0.4431 \quad (9)$$

$$\mu_{no2} = -0.5816, \sigma_{no2} = 1.0082 \quad (10)$$

C - Predict Classification

Note:- Test Features Standardized before computing pdf

$$P(A = yes|Chars = 242, A.W.L = 4.56) = P(A = yes).p(Char s = 242|N(\mu_{yes1}, \sigma_{yes1})).p(Char s = 4.56|N(\mu_{yes2}, \sigma_{yes2}))$$

$$P(A = yes|Chars = 242, A.W.L = 4.56) = 0.6 * 0.2312 * 0.4141 = 0.0574 \quad (11)$$

$$P(A = no|Chars = 242, A.W.L = 4.56) = P(A = no).p(Char s = 242|N(\mu_{no1}, \sigma_{no1})).p(Char s = 4.56|N(\mu_{no2}, \sigma_{no2}))$$

$$P(A = no|Chars = 242, A.W.L = 4.56) = 0.4 * 0.2348 * 0.2457 = 0.0231 \quad (12)$$

$$P(A = yes|Chars = 242, A.W.L = 4.56)_{normalized} = 0.7130 \quad (13)$$

$$P(A = no|Chars = 242, A.W.L = 4.56)_{normalized} = 0.2870 \quad (14)$$

Therefore, Classification = Class - Yes

4:

A - VALIDATION SET

We could use an iterative approach where we use the training data as the model and test it on the Validation Set. We could set $k = 1$. Then perform KNN on the Validation Set with value k . We expect that the overall error on the Validation Set should decrease with $k = 1, 2, \dots$ etc and then reach a locally optimal point after which the error starts increasing which is a good estimate the user-defined parameter k for the model. This k -value may not be the globally optimum k value depending on the nature of the data and do not want to find this global optimum either since we do not want to overfit the validation set.

B - TRAINING SET

KNN does not have an explicit model, the data itself is the model. Therefore, if we use the data(model) to test the model, we would definitely be overfitting the data and the k would not probably not generalize to the unseen test data.

C - TEST SET

If we use the Test Data in KNN to determine the parameter k then it would probably fit the test data well but it would not be an indicator of the true performance of the system. Therefore, we have a separate validation set and the test data gives an estimate of the true underlying performance of the system on unseen data.

5:

$$k(x, y) = \sum_{d=1}^{D=3} x_d y_d = x_1 y_1 + x_2 y_2 + x_3 y_3 = [x_1 \ x_2 \ x_3][y_1 \ y_2 \ y_3]^T = \phi(x)\phi(y)^T$$
$$\phi(x) = [x_1 \ x_2 \ x_3], \phi(y) = [y_1 \ y_2 \ y_3] \quad (15)$$

$$\phi(u) = [u_1 \ u_2 \ u_3] \quad (16)$$

6:

False - This is because the kernel trick is to implicitly compute as if in higher dimensional space but since there are many features and less training data therefore there is less coverage and it makes no sense to compute as if in higher space. Here, the linear kernel should suffice.

PART-2 Naive Bayes Classifier

Precision = 0.6817

Recall = 0.9557

F - measure = 0.7957

Accuracy = 0.8120

PART-3 Multi-Class Support Vector Machines

Accuracy for ONE VS ALL: 0.8812

Accuracy for ONE VS ONE: 0.9024