# HOMEWORK-1

Given Data:

nb\_error:

[0.3591 0.2983 0.2788 0.3087 0.3667 0.3585 0.3784 0.4786 0.3422 0.2135]

Mean:0.26783999999999997

Stdv:0.0667427868761861

dt\_error:

[0.2476 0.1306 0.1197 0.0898 0.2242 0.1946 0.3776 0.2315 0.0062 0.2476]

Mean:0.18694000000000002

Stdv:0.0667427868761861

nn\_error:

[0.2836 0.4117 0.1519 0.3175 0.2301 0.1521 0.2988 0.5041 0.0741 0.2545]

Mean:0.18694000000000002

Stdv:0.0667427868761861

**Q1** Use ANOVA to determine if the three classifiers have equal error rates. Use alpha=0.05.

Ans: **NO**

calculate the SSb = K.sum((means- m)^2)

SSb:0.0436

calculate SSw

nb\_var = np.sum((nb\_error - nb\_mean)\*\*2)

dt\_var = np.sum((dt\_error - dt\_mean)\*\*2)

nn\_var = np.sum((nn\_error - nn\_mean)\*\*2)

SSw = np.sum([nb\_var, dt\_var, nn\_var])

SSw:0.4027

calculate the statistic stat = (SSb/(L-1))/(SSw/l(K-1)) , K = 10, L = 3

Astat = (SSb/2)/(SSw/18)

Astat:0.9750995017735048

from F-table F(0.5,2,18) = 3.5546

Since Astat > F(0.5,2,18) , Hence we reject the H0

**Q2a)** Use Cross-Validated Paired t-test to determine if NB and DecTree have equal errors.

Ans: **NO**

alpha/2 = 0.025

H0: nb\_dt\_mean = 0

nb\_dt = nb\_error - dt\_error

Difference of the error rate per fold:

[ 0.1115 0.1677 0.1591 0.2189 0.1425 0.1639 0.0008 0.2471 0.336

-0.0341]

calculate the t\_stat for the models using the formula sqrt(K)\*(m)/S

t\_stat:4.625

# From t-table, t(apha/2,K-1)

# t(0.025,9) = 2.262

# To accept H0, t\_stat has to be between (-2.262, 2.262)

# t\_stat = 4.6258

# since t\_stat is not in the expected range, we reject H0

**Q2b)** Use Cross-Validated Paired t-test to determine if DecTree and Knearest Neighbor have equal errors.

Ans: **YES**

alpha/2 = 0.025

H0: dt\_nn\_mean = 0

dt\_nn = dt\_error - nn\_error

Difference of the error rate per fold:

[-0.036 -0.2811 -0.0322 -0.2277 -0.0059 0.0425 0.0788 -0.2726 -0.0679

-0.0069]

calculate the t\_stat for the models using the formula sqrt(K)\*(m)/S

t\_stat:-2.056

# From t-table, t(apha/2,K-1)

# t(0.025,9) = 2.262

# To accept H0, t\_stat has to be between (-2.262, 2.262)

# t\_stat = -2.056

# since t\_stat is within the expected range, we accept H0

**Q3)** For each classifier (Naive Bayes, Decision Tree, Knearest Neighbor), determine if the error of the classifier less than p0 (0.3) with level of significance (alpha) (=0.10, 0.05, 0.025).

Ans: **Yes**

p0 = 0.3

H0: mu < 0.3

calculate the t\_stat for the models using the formula sqrt(K)\*(m - p0)/S

nb\_stat:-1.523

dt\_stat:-5.356

nn\_stat:-5.356

###from t-table, t(alpha,K-1), Here DF= K-1 = 9

##for alpha 0.10,and DF =9, tc= 1.383

# for NB: nb\_stat = -1.523 < tc , Hence we accept the null hypothesis

#for DT: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis

#for KNN: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis

##for alpha 0.05,and DF =9, tc=1.833

# for NB: nb\_stat = -1.523 < tc , Hence we accept the null hypothesis

#for DT: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis

#for KNN: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis

#for alpha 0.025,and DF =9, tc=2.262

# for NB: nb\_stat = -1.523 < tc , Hence we accept the null hypothesis

#for DT: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis

#for KNN: nb\_stat = -5.356 < tc , Hence we accept the null hypothesis