Aprendizagem 2023/24

Homework II – Group 029

(ist1103000, ist1103179)

I. Pen-and-paper

1

(a) In keeping with the aforementioned assumptions, $\boxed{y_1 \times y_2}$ follows a multivariate gaussian distribution whose conditional parameters (i.e. in regard to classes A and B) are:

μ_A, μ_B, Σ_A and Σ_B

$$\mu_A = \frac{1}{3}\left(\begin{bmatrix}0.24\\0.36\end{bmatrix}+\begin{bmatrix}0.16\\0.48\end{bmatrix}+\begin{bmatrix}0.32\\0.72\end{bmatrix}\right)=\begin{bmatrix}0.24\\0.52\end{bmatrix} \qquad \mu_B = \overbrace{\frac{1}{4}\left(\begin{bmatrix}0.54\\0.11\end{bmatrix}+\begin{bmatrix}0.66\\0.39\end{bmatrix}+\begin{bmatrix}0.76\\0.28\end{bmatrix}+\begin{bmatrix}0.41\\0.53\end{bmatrix}\right)}^{\frac{1}{N}\sum x_i}=\begin{bmatrix}0.5925\\0.3275\end{bmatrix}$$

$$\Sigma_A = \underbrace{\frac{1}{2}\left(\begin{bmatrix}0\\-0.16\end{bmatrix}\begin{bmatrix}0 & -0.16\end{bmatrix}+\begin{bmatrix}-0.08\\-0.04\end{bmatrix}\begin{bmatrix}-0.08 & -0.04\end{bmatrix}+\begin{bmatrix}0.08\\0.2\end{bmatrix}\begin{bmatrix}0.08 & 0.2\end{bmatrix}\right)}_{\frac{1}{N-1}\sum (x_i-\mu)(x_i-\mu)^T}=\begin{bmatrix}6.4\times10^{-3} & 9.6\times10^{-3}\\9.6\times10^{-3} & 0.0336\end{bmatrix}$$

$$\Sigma_B = \frac{1}{3}\left(\begin{bmatrix}-0.0525\\-0.2175\end{bmatrix}\begin{bmatrix}-0.0525 & -0.2175\end{bmatrix}+\begin{bmatrix}0.0675\\0.0625\end{bmatrix}\begin{bmatrix}0.0675 & 0.0625\end{bmatrix}+\begin{bmatrix}0.1675\\-0.0475\end{bmatrix}\begin{bmatrix}0.1675 & -0.0475\end{bmatrix}+\right.$$

$$\left.+\begin{bmatrix}-0.1825\\0.2025\end{bmatrix}\begin{bmatrix}-0.1825 & 0.2025\end{bmatrix}\right)=\begin{bmatrix}0.023 & -9.912\times10^{-3}\\-9.912\times10^{-3} & 0.031\end{bmatrix}$$

As for the other sets of variables, we only need to compute the conditional probabilities regarding classes A and B:

↳ $\boxed{y_3 \text{ and } y_4}$

- $P(y_3=0, y_4=0)=\frac{2}{7}$; $P(y_3=1,y_4=0)=\frac{2}{7}$; $P(y_3=0,y_4=1)=\frac{2}{7}$; $P(y_3=1,y_4=1)=\frac{1}{7}$
- Taking these values into consideration, we can now compute the conditional probabilities:

$$P(y_3=0,y_4=0\mid y_6=A) \underset{\substack{\downarrow\\ \text{Bayes'}\\ \text{theorem}}}{=} \frac{P(y_6=A\mid y_3=0,y_4=0)\cdot P(y_3=0,y_4=0)}{P(y_6=A)} = \frac{0\times\frac{2}{7}}{\frac{3}{7}} = 0 \ ; \ P(y_3=1,y_4=0\mid y_6=A)=\frac{1}{3}$$

$P(y_3=1,y_4=1\mid y_6=A)=\frac{1}{3}$; $P(y_3=0,y_4=1\mid y_6=A)=\frac{1}{3}$ ⇒ Class A

$P(y_3=0,y_4=0\mid y_6=B)=\frac{1}{2}$; $P(y_3=1,y_4=0\mid y_6=B)=\frac{1}{4}$; $P(y_3=1,y_4=1\mid y_6=B)=0$; $P(y_3=0,y_4=1\mid y_6=B)=\frac{1}{4}$

↳ Class B

↳ $\boxed{y_5}$

- $P(y_5=0)=\frac{2}{7}$; $P(y_5=1)=\frac{3}{7}$; $P(y_5=2)=\frac{2}{7}$
- Taking these values into consideration, we can now compute the conditional probabilities:

$$P(y_5=0\mid y_6=A) \underset{\substack{\downarrow\\ \text{Bayes'}\\ \text{theorem}}}{=} \frac{P(y_6=A\mid y_5=0)\cdot P(y_5=0)}{P(y_6=A)} = \frac{\frac{1}{2}\times\frac{2}{7}}{\frac{3}{7}} = \frac{1}{3} \ ; \ P(y_5=1\mid y_6=A)=\frac{1}{3} \ ; \ P(y_5=2\mid y_6=A)=\frac{1}{3}$$

↳ Class A

$P(y_5=0\mid y_6=B)=\frac{1}{4}$; $P(y_5=1\mid y_6=B)=\frac{1}{2}$; $P(y_5=2\mid y_6=B)=\frac{1}{4}$ ⇒ Class B

With respect to the priors:

- $P(y_6=A)=\frac{3}{7}$
- $P(y_6=B)=\frac{4}{7}$

(b) Under a Maximum A Posteriori (MAP) assumption

$$\hat{y}_6 = \operatorname{argmax}_{c_i} \frac{P(x\mid c_i)\cdot P(c_i)}{P(x)} \Rightarrow \operatorname{argmax}_{c_i} P(x\mid c_i)\cdot P(c_i)$$

↳ $\boxed{\text{Classification of } x_8}$

When $c_i = A$:

$P(x_8\mid A)\cdot P(A) \rightarrow \frac{3}{7}$

$$\hookrightarrow P(y_1=0.38;\ y_2=0.52\mid y_6=A)\cdot \underbrace{P(y_3=0,y_4=1\mid y_6=A)}_{\frac{1}{3}}\cdot \underbrace{P(y_5=0\mid y_6=A)}_{\frac{1}{3}}$$

$P(y_1=0.38;\, y_2=0.52 \mid y_6=A) = \left(1/2\pi\cdot\sqrt{|\Sigma|}\right)\cdot \exp\left(-\frac{1}{2}(x-\mu)^T\cdot\Sigma^{-1}\cdot(x-\mu)\right)$ *

$\Sigma = \Sigma_A$
$\mu = \mu_A$

$\hookrightarrow N\left(\mu_A=\begin{bmatrix}0.24\\0.52\end{bmatrix},\ \Sigma_A=\begin{bmatrix}6.4\times10^{-3} & 9.6\times10^{-3}\\ 9.6\times10^{-3} & 0.0336\end{bmatrix}\right)$

$$\Sigma_A^{-1} = \frac{1}{|\Sigma_A|}\begin{bmatrix}0.0336 & -9.6\times10^{-3}\\ -9.6\times10^{-3} & 6.4\times10^{-3}\end{bmatrix} = \begin{bmatrix}273.4375 & -78.125\\ -78.125 & 52.08(3)\end{bmatrix}$$

$6.4\times10^{-3}\cdot 0.0336 - 9.6\cdot10^{-3}\cdot 9.6\cdot10^{-3} \simeq 0.000123$

$\sqrt{|\Sigma_A|} \simeq 0.011 \qquad \begin{bmatrix}0.38\\0.52\end{bmatrix}-\begin{bmatrix}0.24\\0.52\end{bmatrix}=\begin{bmatrix}0.14\\0\end{bmatrix}$

$$* = \frac{1}{2\pi\times0.011}\exp\left(-\frac{1}{2}\begin{bmatrix}0.14 & 0\end{bmatrix}\begin{bmatrix}273.4375 & -78.125\\ -78.125 & 52.08(3)\end{bmatrix}\begin{bmatrix}0.14\\0\end{bmatrix}\right) \simeq 0.992$$

and thus, $P(x_8|A)\cdot P(A) = 0.992\times\frac{1}{3}\times\frac{1}{3}\times\frac{3}{7} \simeq 0.047$

when $c_i = B$:

$P(x_8|B)\cdot P(B) \rightarrow \frac{4}{7}$

$\hookrightarrow P(y_1=0.38;\, y_2=0.52 \mid y_6=B)\cdot \underbrace{P(y_3=0,\, y_4=1 \mid y_6=B)}_{1/4}\cdot \underbrace{P(y_5=0 \mid y_6=B)}_{1/4}$

$$P(y_1=0.38;\, y_2=0.52 \mid y_6=B) = \frac{1}{2\pi\times0.0248}\exp\left(-1/2\begin{bmatrix}-0.2125 & 0.1925\end{bmatrix}\begin{bmatrix}50.427 & 16.124\\ 16.124 & 37.413\end{bmatrix}\begin{bmatrix}-0.2125\\0.1925\end{bmatrix}\right)\ **$$

$\Sigma = \Sigma_B$
$\mu = \mu_B$

$\hookrightarrow N\left(\mu_B=\begin{bmatrix}0.5925\\0.3275\end{bmatrix},\ \Sigma_B=\begin{bmatrix}0.023 & -9.912\times10^{-3}\\ -9.912\times10^{-3} & 0.031\end{bmatrix}\right)$

$\Sigma_B^{-1} = \begin{bmatrix}50.427 & 16.124\\ 16.124 & 37.413\end{bmatrix} \quad |\Sigma_B| = 0.000615 \quad \sqrt{|\Sigma_B|}\simeq 0.0248 \quad \begin{bmatrix}0.38\\0.52\end{bmatrix}-\begin{bmatrix}0.5925\\0.3275\end{bmatrix}=\begin{bmatrix}-0.2125\\0.1925\end{bmatrix}$

$** \simeq 1.987$

and thus, $P(x_8|B)\cdot P(B) = 1.987\times\frac{1}{4}\times\frac{1}{4}\times\frac{4}{7}\simeq 0.0710$

since $P(x_8|B)\cdot P(B) > P(x_8|A)\cdot P(A) \Leftrightarrow 0.0710 > 0.047$, we can conclude that x_8 is classified as B.

$\hookrightarrow$ Classification of x_9

when $c_i = A$:

$P(x_9|A)\cdot P(A) \rightarrow \frac{3}{7}$

$\hookrightarrow P(y_1=0.42;\, y_2=0.59 \mid y_6=A)\cdot \underbrace{P(y_3=0,\, y_4=1 \mid y_6=A)}_{1/3}\cdot \underbrace{P(y_5=1 \mid y_6=A)}_{1/3}$

$$P(y_1=0.42;\, y_2=0.59 \mid y_6=A) = \frac{1}{2\pi\times0.011}\exp\left(-\frac{1}{2}\begin{bmatrix}0.18 & 0.07\end{bmatrix}\begin{bmatrix}273.4375 & -78.125\\ -78.125 & 52.08(3)\end{bmatrix}\begin{bmatrix}0.18\\0.07\end{bmatrix}\right) \simeq$$

$\begin{bmatrix}0.42\\0.59\end{bmatrix}-\begin{bmatrix}0.24\\0.52\end{bmatrix} = \begin{bmatrix}0.18\\0.07\end{bmatrix}$

$\simeq 0.406$

and thus $P(x_9|A)\cdot P(A) = 0.406\times\frac{1}{3}\times\frac{1}{3}\times\frac{3}{7}\simeq 0.019$

when $c_i = B$:

$P(x_9|B)\cdot P(B) \rightarrow \frac{4}{7}$

$\hookrightarrow P(y_1=0.42;\, y_2=0.59 \mid y_6=B)\cdot \underbrace{P(y_3=0,\, y_4=1 \mid y_6=B)}_{1/4}\cdot \underbrace{P(y_5=1 \mid y_6=B)}_{1/2}$

$\begin{bmatrix}0.42\\0.59\end{bmatrix}-\begin{bmatrix}0.5925\\0.3275\end{bmatrix} = \begin{bmatrix}-0.1725\\0.2625\end{bmatrix}$

$$P(y_1=0.42;\, y_2=0.59 \mid y_6=B) = \frac{1}{2\pi\times0.0248}\exp\left(-1/2\begin{bmatrix}-0.1725 & 0.2625\end{bmatrix}\begin{bmatrix}50.427 & 16.124\\ 16.124 & 37.413\end{bmatrix}\begin{bmatrix}-0.1725\\0.2625\end{bmatrix}\right)$$

$\simeq 1.733$

and thus $P(x_9|B)\cdot P(B) = 1.733\times\frac{1}{4}\times\frac{1}{2}\times\frac{4}{7}\simeq 0.124$

Since $P(x_9|B)\cdot P(B) > P(x_9|A)\cdot P(A) \Leftrightarrow 0.124 > 0.019$, we can conclude that x_9 is classified as B.

(c) under a Maximum Likelihood (ML) assumption

$\hat{y}_6 = \arg\max_{c_i} P(x|c_i)$

and thus:

$x_8 \rightarrow P(x_8|A) = 0.992 \times \frac{1}{3} \times \frac{1}{3} = 0.110 \leftarrow$ observation with $y_6 = A$ with lowest $P(x_8|A)$

$x_9 \rightarrow P(x_9|A) = 0.406 \times 1/3 \times 1/3 = 0.045 \leftarrow$ observation with $y_6 = B$ with lowest $P(x_9|A)$

normalization

$P(x_8|A) = \frac{0.110}{0.110 + \underbrace{0.124}_{P(x_8|B)}} \simeq 0.470$

$P(x_9|A) = \frac{0.045}{\underbrace{0.217}_{P(x_9|B)} + 0.045} \simeq 0.172$

In conclusion, in order to optimize testing accuracy, $\theta \in]0{,}172 ; 0{,}470[$.

2

(a)

$\overline{y_2} = 0.4422$

$\left.\begin{array}{l}\min y_2 = 0.11 \\ \max y_2 = 0.72\end{array}\right\}$ range $y_2 = 1 - 0 = 1 \Rightarrow$ range $= \frac{1}{2} = 0.5$

$\left.\begin{array}{l}\hat{y}_2\text{ new} = 0 \Rightarrow y_2 \in [0; 0.5[\\ \hat{y}_2\text{ new} = 1 \Rightarrow y_2 \in [0.5; 1]\end{array}\right\} y_2 = [0\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1]$

Folds:

↳ #1

$x_1 \rightarrow y_2 = 0, y_3 = 1, y_4 = 1, y_5 = 0, y_6 = A$

$x_2 \rightarrow y_2 = 0, y_3 = 1, y_4 = 0, y_5 = 1, y_6 = A$

$x_3 \rightarrow y_2 = 1, y_3 = 0, y_4 = 1, y_5 = 2, y_6 = A$

↳ #2

$x_4 \rightarrow y_2 = 0, y_3 = 0, y_4 = 0, y_5 = 1, y_6 = B$

$x_5 \rightarrow y_2 = 0, y_3 = 0, y_4 = 0, y_5 = 0, y_6 = B$

$x_6 \rightarrow y_2 = 0, y_3 = 1, y_4 = 0, y_5 = 2, y_6 = B$

↳ #3

$x_7 \rightarrow y_2 = 1, y_3 = 0, y_4 = 1, y_5 = 1, y_6 = B$

$x_8 \rightarrow y_2 = 1, y_3 = 0, y_4 = 1, y_5 = 0, y_6 = A$

$x_9 \rightarrow y_2 = 1, y_3 = 0, y_4 = 1, y_5 = 1, y_6 = B$

(b) training observations: $x_1 - x_6$ (folds 1 and 2)

testing observations: $x_7 - x_9$ (fold 3)

Hammington distance: $d(x_i, x_j) = \sum_{k=1}^{p} y_{ki} \neq y_{kj}$ $\quad w_i = \frac{1}{d(x, x_i)}$ $\quad w_i' = \frac{w_i}{\sum w_i}$

x_7:

$d(x_7, x_1) = 4$

$d(x_7, x_2) = 4$

$d(x_7, x_3) = 2$

$d(x_7, x_4) = 2$

$d(x_7, x_5) = 3$

$d(x_7, x_6) = 4$

$\rightarrow$ nearest neighbours: x_3, x_4, x_5 $\quad \sum w_i = \frac{1}{2} + \frac{1}{2} + \frac{1}{3} = \frac{4}{3}$

$\hat{y}_1(x_7) = \frac{1/2}{4/3} \times 0.32 + \frac{1/2}{4/3} \times 0.54 + \frac{1/3}{4/3} \times 0.66 = 0.4875$

x_8:

$d(x_8, x_1) = 2$

$d(x_8, x_2) = 4$

$d(x_8, x_3) = 1$

$d(x_8, x_4) = 4$

$d(x_8, x_5) = 3$

$d(x_8, x_6) = 5$

$\rightarrow$ nearest neighbours: x_1, x_3, x_5 $\quad \sum w_i = \frac{1}{2} + 1 + \frac{1}{3} = \frac{11}{6}$

$\hat{y}_1(x_8) = \frac{1/2}{11/6} \times 0.24 + \frac{1}{11/6} \times 0.32 + \frac{1/3}{11/6} \times 0.66 = 0.36$

x_9:

$d(x_9, x_1) = 4$

$d(x_9, x_2) = 4$

$d(x_9, x_3) = 2$

$d(x_9, x_4) = 2$

$d(x_9, x_5) = 3$

$d(x_9, x_6) = 4$

$\rightarrow$ nearest neighbours: x_3, x_4, x_5 $\quad \sum w_i = 4/3$

$\hat{y}_1(x_9) = \frac{1/2}{4/3} \times 0.32 + \frac{1/2}{4/3} \times 0.54 + \frac{1/3}{4/3} \times 0.66 = 0.4875$

$MAE = \frac{1}{n} \sum_{i=1}^{n} |z_i - \hat{z}_i| = \frac{|0.41 - 0.4875| \times 2 + |0.38 - 0.36|}{3} = 0.05833$

Aprendizagem 2023/24

Homework II – Group 029

(ist1103000, ist1103179)

II. Programming and critical analysis

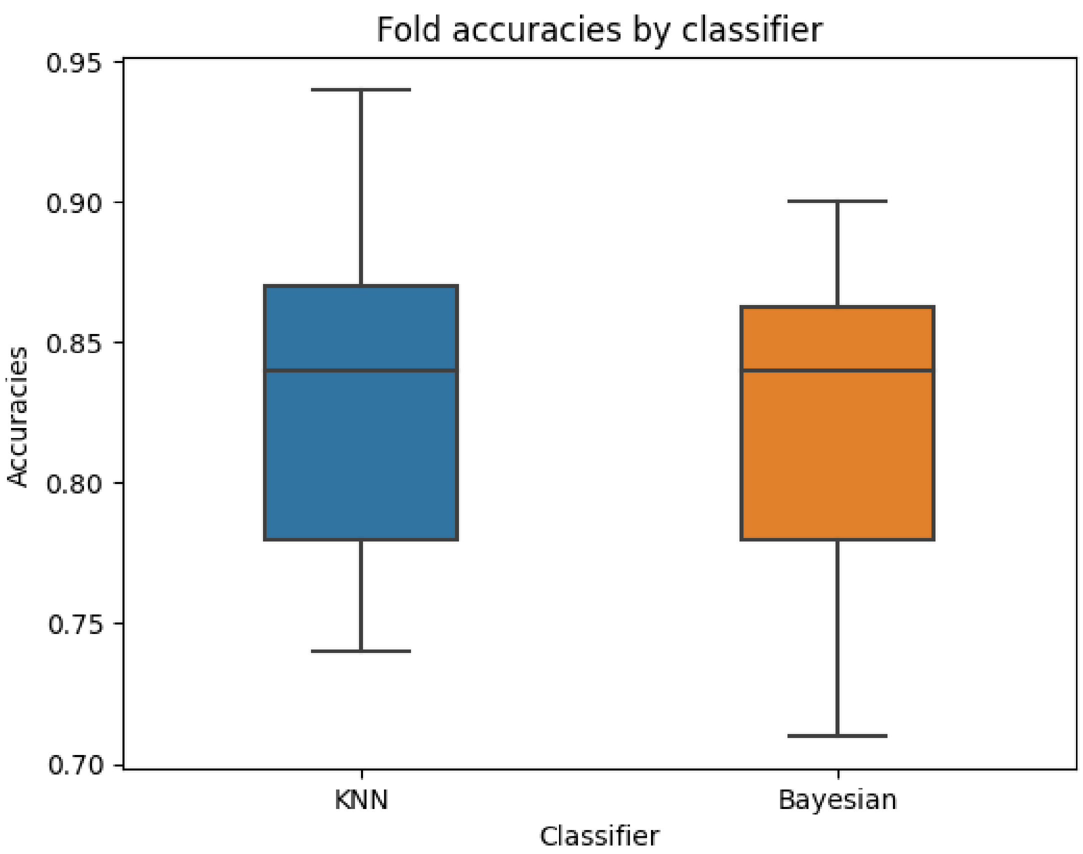
1)

a) Code and graphs:

```
import pandas as pd
from scipy.io.arff import loadarff

data = loadarff('column_diagnosis.arff')
df = pd.DataFrame(data[0])
df['class'] = df['class'].str.decode('utf-8')
df.head()


from sklearn import metrics
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB

from sklearn.model_selection import StratifiedKFold

input_vars = df.drop("class", axis=1)
output_vars = df["class"]

knn_classifier = KNeighborsClassifier(n_neighbors=5)
naive_bayes_calssifier = GaussianNB()

folds = StratifiedKFold(n_splits=10, shuffle=True, random_state=0)
knn_fold_accuracies = []
naive_bayes_fold_accuracies = []
for train_k, test_k in folds.split(input_vars, output_vars):
    X_train, X_test = input_vars.iloc[train_k], input_vars.iloc[test_k]
    y_train, y_test = output_vars.iloc[train_k], output_vars.iloc[test_k]

    knn_classifier.fit(X_train, y_train)
    y_pred = knn_classifier.predict(X_test)
    knn_fold_accuracies.append(round(metrics.accuracy_score(y_test,
y_pred),2))

    naive_bayes_calssifier.fit(X_train, y_train)
    y_pred = naive_bayes_calssifier.predict(X_test)
    naive_bayes_fold_accuracies.append(round(metrics.accuracy_score(y_test,
y_pred),2))

print(knn_fold_accuracies)
print(naive_bayes_fold_accuracies)

[0.94, 0.81, 0.87, 0.94, 0.74, 0.87, 0.84, 0.84, 0.77, 0.77]
```

```
[0.84, 0.87, 0.84, 0.87, 0.77, 0.84, 0.9, 0.81, 0.77, 0.71]
```

```
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.DataFrame({'Accuracies':
knn_fold_accuracies+naive_bayes_fold_accuracies, 'Classifier':
["KNN"]*10+["Bayesian"]*10})
sns.boxplot(data=df, x='Classifier', y='Accuracies',
width=0.4).set(title='Fold accuracies by classifier')
plt.show()
```



b) Code and comments:

```
from scipy.stats import ttest_rel

test_results = ttest_rel(knn_fold_accuracies, naive_bayes_fold_accuracies,
alternative='less')
print(test_results)
```

```
TtestResult(statistic=0.9923982173934255, pvalue=0.8265333762138204, df=9)
```

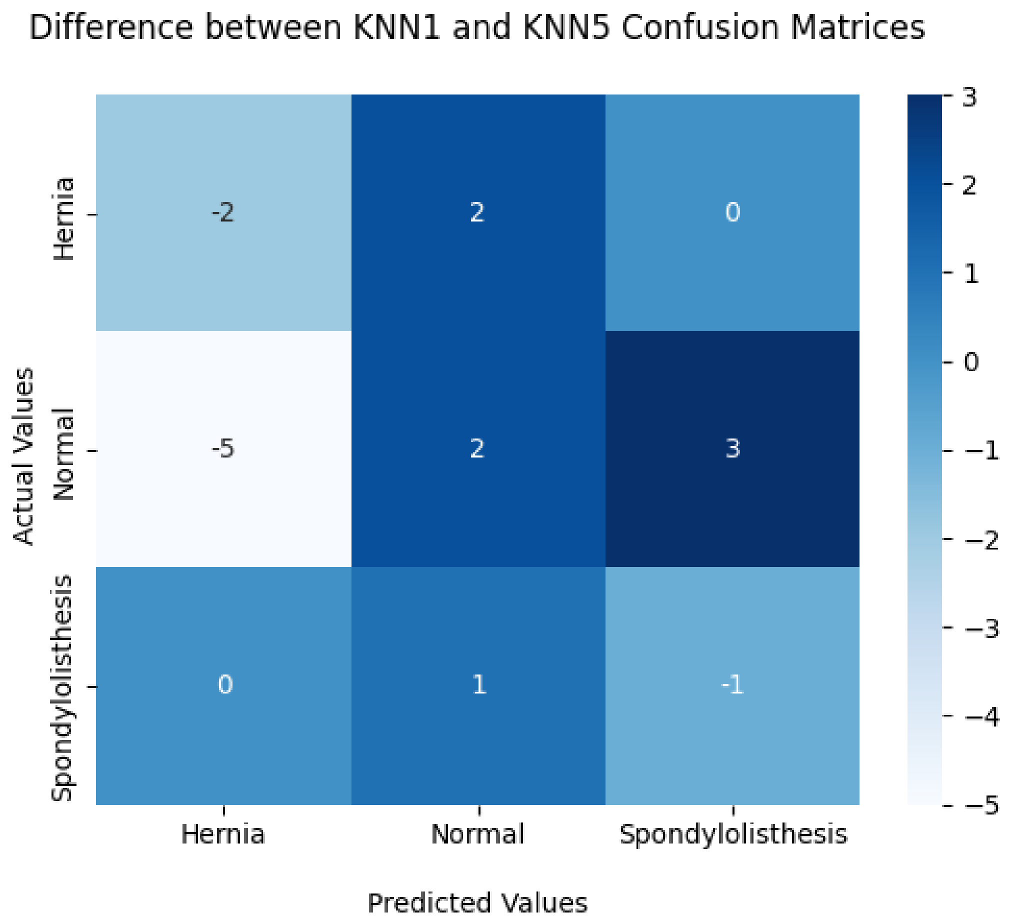
Since our alternative hypothesis is the KNN accuracy mean being less than the Bayesian accuracy mean and the value obtained is about 82,70%, we can't reject the alternative hypothesis for standard significance levels, so we can't certainly assume that the null hypothesis is true.

2) Code, graphs and comments:

```
from sklearn.metrics import confusion_matrix
from functools import reduce

knn_classifier_1 = KNeighborsClassifier(n_neighbors=1, metric="euclidean")
knn_classifier_5 = KNeighborsClassifier(n_neighbors=5, metric="euclidean")

confusion_matrix_knn1 = []
confusion_matrix_knn5 = []
for train_k, test_k in folds.split(input_vars, output_vars):
    X_train, X_test = input_vars.iloc[train_k], input_vars.iloc[test_k]
    y_train, y_test = output_vars.iloc[train_k], output_vars.iloc[test_k]

    knn_classifier_1.fit(X_train, y_train)
    y_pred = knn_classifier_1.predict(X_test)
    cm = confusion_matrix(y_test, y_pred, labels=["Hernia" , "Normal",
"Spondylolisthesis"])
    confusion_matrix_knn1.append(cm)

    knn_classifier_5.fit(X_train, y_train)
    y_pred = knn_classifier_5.predict(X_test)
    cm = confusion_matrix(y_test, y_pred, labels=["Hernia" , "Normal",
"Spondylolisthesis"])
    confusion_matrix_knn5.append(cm)

cumulative_cm_knn1 = (reduce(lambda a,b: a+b, confusion_matrix_knn1))
cumulative_cm_knn5 = (reduce(lambda a,b: a+b, confusion_matrix_knn5))

print(cumulative_cm_knn1)
print(cumulative_cm_knn5)

diff_cm = cumulative_cm_knn1 - cumulative_cm_knn5

ax = sns.heatmap(diff_cm, annot=True, cmap='Blues')

ax.set_title('Difference between KNN1 and KNN5 Confusion Matrices\n')
ax.set_xlabel('\nPredicted Values')
ax.set_ylabel('Actual Values ')

## Ticket labels - List must be in alphabetical order
```

```
ax.xaxis.set_ticklabels(["Hernia","Normal","Spondylolisthesis"])
ax.yaxis.set_ticklabels(["Hernia","Normal","Spondylolisthesis"])

## Display the visualization of the Confusion Matrix.
plt.show()
```

```
[[ 37  23   0]
 [ 14  80   6]
 [  1   7 142]]
[[ 39  21   0]
 [ 19  78   3]
 [  1   6 143]]
```



Comment:

As a result of our analysis of the previous plot, it can be deduced that there isn't much of a difference between the observations registered using KNN1 and KNN5 (it goes neither under -5 nor more than 2) and thus both approaches end up having similar behaviour.

3)

When learning a naive Bayes model from the column diagnosis dataset, we assume independence among every input variable, which may or may not be true, because we don't know how these variables are related and this could lead to incorrect results (e.g., there are three features related to "pelvis", so they're more likely to be associated with each other than not). Additionally, our dataset only has continuous features and thus our results may not be optimal as some of these features may not follow a Gaussian distribution — an assumption made by Naïve Bayes. This model is also affected by the presence of extraneous data and/or variables and may give them undue importance when making predictions.

END