

1. Application of Gaussian Naive Bayes

In this part you may change the packages/modules as you wish. The structure listed below is just to give you an idea of what is expected.

```
1 import numpy as np
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 from sklearn import datasets
6 from sklearn.model_selection import train_test_split
7 from sklearn.metrics import accuracy_score
8 from sklearn.datasets import load_breast_cancer
9 from naive_bayes import GaussianNB
10
11 # Load breast cancer data using load_breast_cancer and inspect it.
12 cancer =
13
14 # Get the data and target
15 X, y =
16
17 # Get feature names
18 feature_names =
19 print(feature_names)
20
21 # Create a pandas dataframe
22 df_cancer =
23
24 # Compute pairwise correlation of features.
25 # See https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html
26 corr =
27
28 # Plot correlations using seaborn or any other plotting tool.
29
30
31
32 # Can you use all of the features? Remember the fundamental assumption of
33 # naive bayes. Explain your thinking.
34
35
36
37 # There are many features. Select some of them (maybe four or five) and get
38 # predictions for your test samples using GaussianNB class that you have
39 # implemented in the previous homework.
40
41
42 # Your selected features
43 X =
44 # Labels of your selected features
45 y =
46
47 # Split the data using train_test_split
48 X_train, X_test, y_train, y_test =
49
50
51 clf = GaussianNB()
52 clf.fit(X_train, y_train)
53 predictions = clf.predict(X_test)
54 print("Accuracy score:", accuracy_score(y_test, predictions))
```