Neelay Chakravarthy

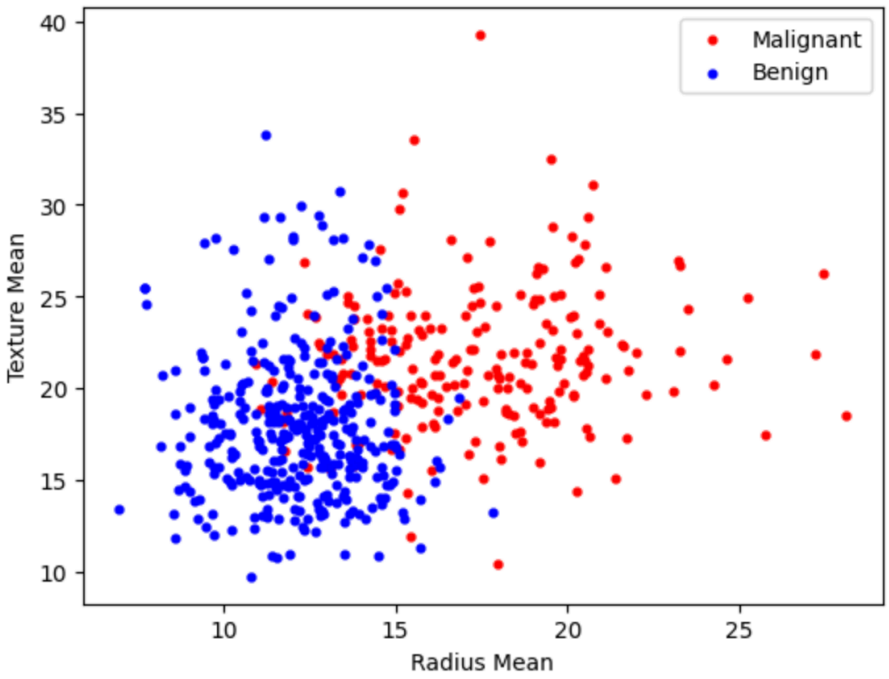
Dr. Chen

CSCI 184

May 15th, 2024

HW2 Part 2 Report

```
            label="Malignant")
plt.scatter(cancer_raw_data["radius_mean"][cancer_raw_data.diagnosis == "B"],
            cancer_raw_data["texture_mean"][cancer_raw_data.diagnosis == "B"],
            c='b',
            s=12,
            marker='o',
            label="Benign")
plt.xlabel("Radius Mean")
plt.ylabel("Texture Mean")
plt.legend()
plt.show()
#Data is not fully linearly separable
```



I decided to use Gaussian Naive Bayes. This is because the features are continuous, and since we are dealing with radius and textures of tumors, the assumption of gaussian distributed features is okay. Here were the results:

```
print(f"Training Recall Score:{recall_score(Y_pred_train
```

```
Training Accuracy: 0.6206030150753769
Training Confusion Matrix: [[244 146]
 [  5   3]]
Training F1 score: 0.03821656050955414
Training Precision Score: 0.020134228187919462
Training Recall Score:0.375
```

```
Testing Accuracy: 0.6374269005847953
Testing Confusion Matrix: [[107  61]
 [  1   2]]
Testing F1 score: 0.0606060606060606
Testing Precision Score: 0.031746031746031744
Testing Recall Score:0.6666666666666666
```

I also tried Multinomial, Complement, and Bernoulli for fun, which are listed in order below:

```
Training Accuracy: 0.4020100502512563
Training Confusion Matrix: [[ 32  21]
 [217 128]]
Training F1 score: 0.5182186234817815
Training Precision Score: 0.8590604026845637
Training Recall Score:0.3710144927536232
Testing Accuracy: 0.3391812865497076
Testing Confusion Matrix: [[11 16]
 [97 47]]
Testing F1 score: 0.4541062801932367
Testing Precision Score: 0.746031746031746
Testing Recall Score:0.3263888888888889
```

```
Training Accuracy: 0.4020100502512563
Training Confusion Matrix: [[ 32  21]
 [217 128]]
Training F1 score: 0.5182186234817815
Training Precision Score: 0.859060402684563 7
Training Recall Score:0.3710144927536232
Testing Accuracy: 0.3391812865497076
Testing Confusion Matrix: [[11 16]
 [97 47]]
Testing F1 score: 0.4541062801932367 3
Testing Precision Score: 0.746031746031746
Testing Recall Score:0.3263888888888889

Training Accuracy: 0.6256281407035176
Training Confusion Matrix: [[249 149]
 [  0   0]]
Training F1 score: 0.0
Training Precision Score: 0.0
Training Recall Score:0.0
Testing Accuracy: 0.631578947368421
Testing Confusion Matrix: [[108  63]
 [  0   0]]
Testing F1 score: 0.0
Testing Precision Score: 0.0
Testing Recall Score:0.0
```