

# # GDA classification of breast cancer using sklearn

```
In [23]: from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
import matplotlib.pyplot as plt
```

```
In [24]: ## Loading the data
dataset = datasets.load_breast_cancer()
```

```
In [25]: X = dataset.data
Y = dataset.target

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.4)
```

```
In [26]: # initializing the Gaussian classifier
gaussian_classifier = GaussianNB()
```

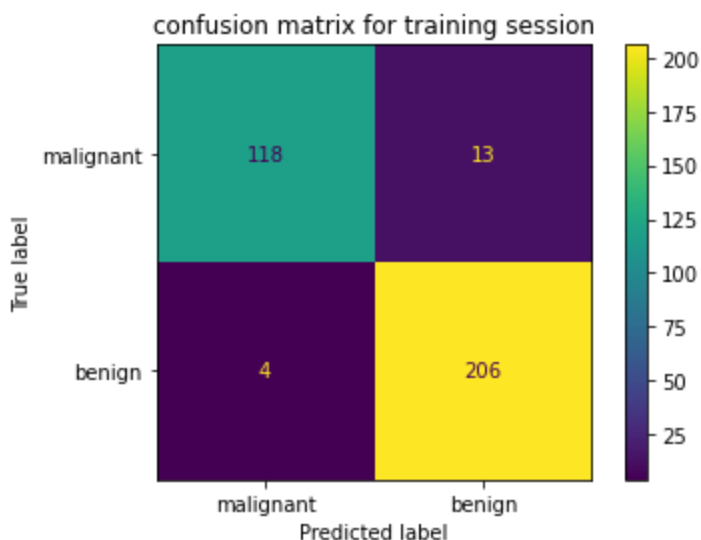
**training the model with 60% of the data, and presenting it**

```
In [27]: # Fit Gaussian Naive Bayes according to X, Y
gaussian_classifier.fit(X_train, Y_train)

# Perform classification on an array of test vectors X
trained_y = gaussian_classifier.predict(X_train)

# creating the matrix to show on graph + naming the classes
confusion_matrix_train = metrics.confusion_matrix(Y_train, trained_y)
names = dataset.target_names

metrics.ConfusionMatrixDisplay(confusion_matrix_train, display_labels=names).plot();
plt.title('confusion matrix for training session');
```



**testing the model with the remaining 40% of the data and presenting it with a confusion matrix**

```
In [28]: # Perform classification on an array of test vectors X
tested_y = gaussian_classifier.predict(X_test)

# creating the matrix to show on graph + naming the classes
confusion_matrix_test = metrics.confusion_matrix(Y_test, tested_y)

metrics.ConfusionMatrixDisplay(confusion_matrix_test, display_labels=names).plot();
plt.title('confusion matrix for testing session');
```

