1. Generate two datasets, X (training set) and X_1 (test set), each consisting of N = 1000 3-dimensional vectors that stem from three classes, $\omega 1$, $\omega 2$, and $\omega 3$, with prior-probabilities P(ω1)=P(ω2)=P(ω3)=1/3. The classes are modeled by Gaussian distributions with means m1 = $[0, 0, 0]^T$, m2 = $[1, 2, 2]^T$, and m3 = $[3, 3, 4]^T$ respectively; their covariance matrices are

$$S_1=\begin{bmatrix}0.8 & 0.2 & 0.1\\ 0.2 & 0.8 & 0.2\\ 0.1 & 0.2 & 0.8\end{bmatrix}, S_2=\begin{bmatrix}0.6 & 0.01 & 0.01\\ 0.01 & 0.8 & 0.01\\ 0.01 & 0.01 & 0.6\end{bmatrix}, S_3=\begin{bmatrix}0.6 & 0.1 & 0.1\\ 0.1 & 0.6 & 0.1\\ 0.1 & 0.1 & 0.6\end{bmatrix}$$

   (a) Use the Euclidean distance classifier to classify the points of X_1.
   (b) Use the Mahalanobis distance classifier to classify the points of X_1.
   (c) Use the Bayesian classifier to classify the points of X_1.
   (d) For each class, compute the error probability and compare the results.
   (e) Experiment with the mean values (bringing them closer or taking them farther away) and the a prior-probabilities. Comment on the results.

2. Considering the California Housing dataset, design a linear regression model considering each feature with non zero values, and report the best feature and model accordng to the R^2 metric.

   (Evaluate your linear regression model using sum of squares due to regression (SSR), sum of squares error (SSE), sum of squares total (SST) and coefficient of determination R^2 metric and adjusted R^2 metric.)

$$SST = SSR + SSE$$

$$\sum_{i=0}^{n}(y_i - \bar{y})^2 = \sum_{i=0}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=0}^{n}(y_i - \hat{y})^2$$

$$R^2 = \frac{SSR}{SST} \qquad Adjusted\ \ R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$$

Note: Use the following code snippet to load the California housing dataset -

```
import sklearn
caldata = sklearn.datasets.fetch_california_housing()
print(caldata.data.shape, caldata.target.shape)
print(caldata.feature_names)
```