

Question 1

TP = 6, TN = 7, FP = 3, FN = 4

Precision: $6/(6+3) = 0.6666666667$

Recall: $6/(6+4) = 0.6$

F-measure: $2 * 0.6666666667 * 0.6 / (0.6666666667 + 0.6) = 0.63157894738$

Accuracy: $(6 + 7) / (6+7+3+4) = 0.65$

Question 2

Storing and Operating on the Data

Firstly some modules are required to be imported such as pandas, math and sklearn. Sklearn gives us the models we are going to use to create a prediction. We use a support vector machine for classification and Linear regression for regression. Sklearn also import the accuracy measurer and the mean squared error to predict the accuracy for the predictions.

Storing the training and testing data required the use of pandas due to their flexibility and the fact that they make data very easy to access. To load the data from the files I use the read.csv method to access the data from the path its given and then put into a DataFrame.

To make the binary labels for expensive and non-expensive houses I call the specific column from the dataframe and then checking if its greater than or equal to 30 and using .astype(int) to change any values greater than or equal to 30 to 1 and changes any values less than 30 equal to 0. To shuffle the dataframe I use .sample(frac = 1) method which randomises the data to prevent overfitting.

To divide the training and testing data into the X and Y datasets i use the .iloc method to specify the correct columns to operate on where X contains all the features that go into predicting and Y contains the column that deals with classifying the house price.

Classification

For the classification model uses a support vector machine which fits the X and Y training sets, then the prediction is done using the X testing set and finally using the Y testing set to show the predictions accuracy which is then displayed on the console using a print statement.

Regression

The regression model utilises a linear regression model to fit the X and Y training sets, then the prediction using the X testing set, calculating the means squared error using the Y testing set against the prediction and then displaying the score on the console after using the .sqrt method from the math module to square root the means squared error to get the root means squared error.