1 Practice

1.1 Classification experiment: Manual precision, recall, f-measure and accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - measure = 2 * \frac{Recall * Precision}{Recall + Precision}$$

$$TP : IDs\{1, 2, 4, 7, 8, 9\}$$

$$TP = 6$$

$$FP : IDs\{13, 18, 19\}$$

$$FP = 3$$

$$TN : IDs\{10, 11, 12, 14, 16, 17, 20\}$$

$$TN = 7$$

$$FN : IDs\{3, 5, 6, 15\}$$

$$FN = 4$$

$$Accuracy = \frac{6 + 7}{6 + 3 + 7 + 4} = \frac{13}{20} = 0.6500$$

$$Precision = \frac{6}{6 + 3} = \frac{2}{3} \approx 0.6667$$

$$Recall = \frac{6}{6 + 4} = \frac{6}{10} = 0.6000$$

$$F - measure = 2 * \frac{0.6 * \frac{2}{3}}{0.6 + \frac{2}{3}} = \frac{12}{19} \approx 0.6316$$

1.2 Real Estate Machine Learning Models

Using a Linear Classification SVM, it was trained using the sklearn svm package where the function SVC (a Classification Support Vector Machine function) was set to have a linear kernel and an automatic gamma, this is initialised as shown below:

```
svm_clf_expensive=sklearn.svm.SVC(kernel="linear", gamma='auto' ,cache_size=2000)

poly_svm = make_pipeline(StandardScaler(), svm_clf_expensive)

poly_svm.fit(X_train_expensive,Y_train_expensive)
```

This gave 72 True Positives, 6 False Positives, 26 True Negatives and 9 False Negatives resulting in an accuracy of

$$Accuracy = \frac{72 + 26}{72 + 6 + 26 + 9} = \frac{98}{113} \approx 0.8673$$

Using a Regression SVM with a polynomial kernel of degree 3 gave a mean squared error of approximately 82.0723.