Report

- **How did you preprocess this dataset ?**
  1. 把 High Price 和 Low Price 取出，求出其中間值 Mid Price，只使用 Mid Price 作為 feature，捨棄 High Price 和 Low Price。
  2. 把 Volume 劃分成五個區間，發現最大的區間的上漲機率最低，因此分成在最大區間(High Volume=1)和其他(High Volume=0)兩種，此為 High Volume 特徵。

  用 Open Price、Close Price、Mid Price、High Volume 四個特徵訓練。

- **Which classifier reaches the highest classification accuracy in this dataset ?**

  Accuracy： Logistic Regression(82%) > Adaboost(55%) > Neural Networks(53%)

  - **Why ?**

  因為直接使用 Volume 特徵時，Logistic Regression 準確度只有五成，但轉換成 High Volume，特徵更加明顯，準確度上升。Adaboost 在較難分的資料上，因為沒有更多更強烈的特徵輔助，所以準確度不高。Neural Network 只使用線性模型，在特徵少的情況下，準確率低。

  - **Can this result remain if the dataset is different ?**

  使用 Google Stock dataset，Logistic Regression 依然是準確度最高，Accuracy： Logistic Regression(75%) > Neural Networks(60%) > Adaboost(45%)。只有 Neural Networks 的準確度上升，其他兩者下降。

- **How did you improve your classifiers ?**
  1. 移除不必要特徵，High Price、Low Price 為極值，變異性較大，且沒有明顯的特徵會影響收盤價。
  2. 新增特徵，把 High Price 和 Low Price 取中間值，此值與 Open Price 和 Close Price 較相近，較值得參考。
  3. 視覺化後觀察特徵並作數值轉換，把 Volume 做數值轉換，並取出漲跌特質較明顯的作為特徵，Logistic Regression 的準確度就大幅上升。
  4. 調整超參數，使用不同的 activation function 使 Neural Networks 的準確度提升。