Report 1

a. kNN (k nearest neighbors)
Used dataset: Iris classification
Iris classification的資料集僅有150筆資料，每筆資料共有4個features及1個target類別共有3類標記為0, 1, 2。實驗以105筆資料作為training data

實驗結果如下圖Fig 1.，橘色線為呼叫sklearn函式庫的執行結果，綠色線為根據演算法實作的執行結果。分別將k從1至100進行測試，可以看出因資料數並不多，在k較小時便有很好的結果，當k值持續增大，會有underfitting的現象，準確度反而大幅下降。（實際執行時間應該要除以100）

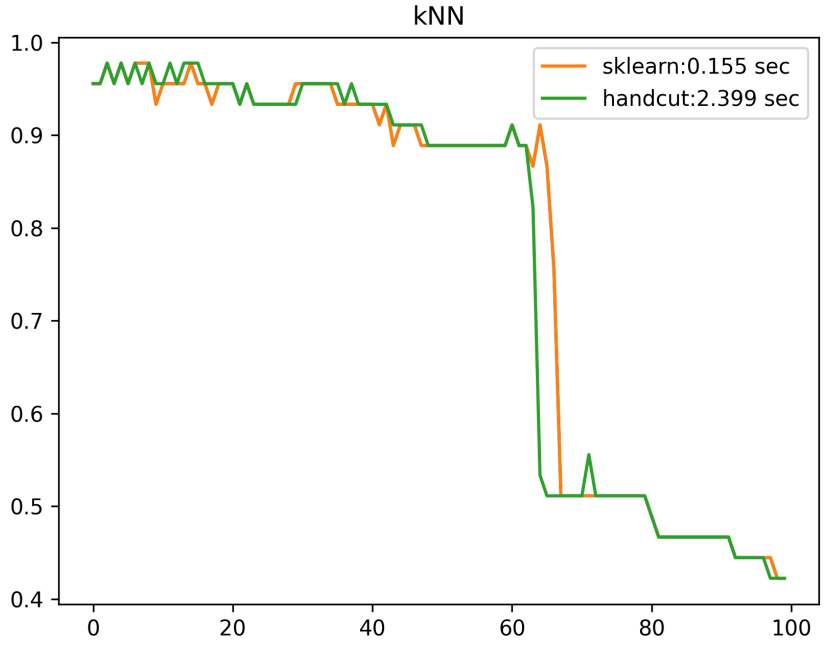

Fig 1

B. Naive Bayes
同樣採用Iris資料集來實驗，觀察到每筆資料的feature都是numerical的，因此決定先將feature的資料進行分組。將每一個feature設定一個threshold，超過門檻值的設定成class 1，反之class 0。因所有feature都已經二值化（二類），套用Bernoulli Naive Bayes 進行實作，結果如Fig 2.

```
sklearn Acc: 0.711
sklearn: 0.001 sec
HandCut Acc: 0.822
HandCut: 0.037 sec
```

Fig 2

C. Logistic Regression
同樣使用Iris資料集，由於是multiclass的問題，所以用One-vs-All的策略，分別訓練3個分類器來各自判斷是否為第一類、是否為第二類、是否為第三類。在比較時便取三者分數最高者作為預測的結果。實驗結果如Fig 3，可以看到Logistic Regression若要讓模型收斂，通常需要比前兩者演算法更長的時間。

```
handcut:0.994 sec
handcut accuracy:0.978
++++++++++++++++++++
sklearn:0.020 sec
sklearn acc:0.978
```

Fig 3