AI&ML選修課的作業

本次作業是使用較為熟悉之程式語言-Matlab進行撰寫,使用貝氏分類器(Bayes classification)進行 鳶尾花的分類,此花卉基本特徵型態為雄蕊三枚,花柱三歧,共六稜,然而鳶尾花種類繁多,本次 專題選擇三個種類的鳶尾花進行分類,分別為setosa、versicolor及virginia,透過貝氏分類器訓練各種 類的樣本,表1為三種鳶尾花的資料範本,共計150張的訓練樣本,其中每一種類各佔50張,接著輸 入貝式分類器進行樣本訓練·本文的目的在於利用鳶尾花花瓣的寬度以及長度進行資料的分類,並 且訓練鳶尾花的特徵利用貝氏分類氣將其簡單區分為三大區塊·

貝氏分類演算法具有以下幾點特點:主要是基於機率來分類(probabilistic classification),使用貝氏定理進行計算,並且假設特徵事件互相獨立(independence),利用以上特點進行樣本數訓練以及近一步的分類:貝氏分類對於少量的訓練集合一樣會有不錯的分類準度,利用所謂的先驗分配,進而補足小樣本的不足,使得推論能夠進行,是其一大特點:

表1鳶尾花資料庫

Value	Count	Percent	Picture
setosa	50	33.33%	Iris setosa
versicolor	50	33.33%	Iris versicolor
virginia	50	33.33%	Irls virginica

```
MATLAB 程式碼及註解如下:
CLEAR;CLC
```

```
%%輸入鳶尾花的資料
  load fisheriris
  X = meas(:,3:4);
  Y = species;
  tabulate(Y)
  %%使用貝氏分類器進行樣本資料的分類
  %訓練樣本並且指定順序
  Mdl = fitcnb(X, Y, ...
     'ClassNames', {'setosa', 'versicolor', 'virginica'})
  % |Mdl| 是訓練的分類.
  % 使用具有一些均值和標準差的高斯分佈對每個分類中的預測分佈進行建模
% 顯示setosa內第一特徵的擬合 EX:seto
  setosaIndex = strcmp(Mdl.ClassNames,'setosa');
  estimates = Mdl.DistributionParameters{setosaIndex,1}
  % 平均值是1.4620; 標準差是0.1737.
  % 畫高斯等高線圖.
  figure
  gscatter(X(:,1),X(:,2),Y);
  h = gca;
  exlim = h.XLim;
  cylim = h.YLim;
  hold on
  Params = cell2mat(Mdl.DistributionParameters);
  Mu = Params(2*(1:3)-1,1:2); \% Extract the means
  Sigma = zeros(2,2,3);
  for i = 1:3
   Sigma(:,:,j) = diag(Params(2*j,:)).^2; % Create diagonal covariance matrix
     xlim = Mu(j,1) + 4*[1-1]*sqrt(Sigma(1,1,j));
    ylim = Mu(j,2) + 4*[1-1]*sqrt(Sigma(2,2,j));
     ezcontour(@(x1,x2)mvnpdf([x1,x2],Mu(j,:),Sigma(:,:,j)),[xlim ylim])
  end
  h.XLim = cxlim:
  h.YLim = cylim;
  title('Naive Bayes Classifier -- Fisher''s Iris Data')
  xlabel('Petal Length (cm)')
  ylabel('Petal Width (cm)')
  hold off
```

結果討論:

下圖為貝氏分類器用於鳶尾花的程式執行結果,橫軸為花瓣長度、縱軸為花瓣寬度,圖中紅色點為setosa、綠色點為versicolor、藍色點為virginia,透過貝氏分類器可以將花辦的特徵進行分類,一共分為三類,可以明顯看出這些樣本經過此分類器後明顯畫出三大塊同心圓data1、data2、及data3,因此,本文認為經過此演算法電腦會自動給予一組特徵向量以及特徵值,協助電腦繪製並分類出三種鳶尾花的種類,未來提供一組測試樣本後,電腦會依據訓練樣本提供的特徵值,協助判斷測試樣本的種類.

