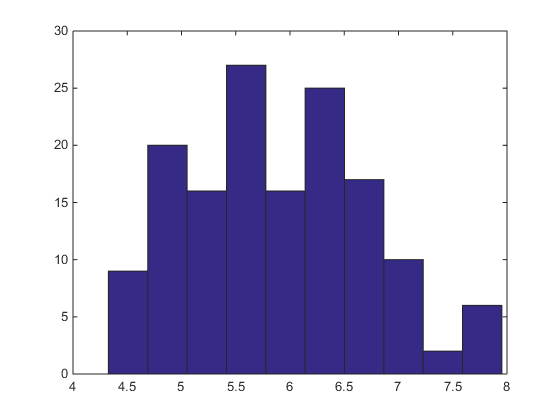
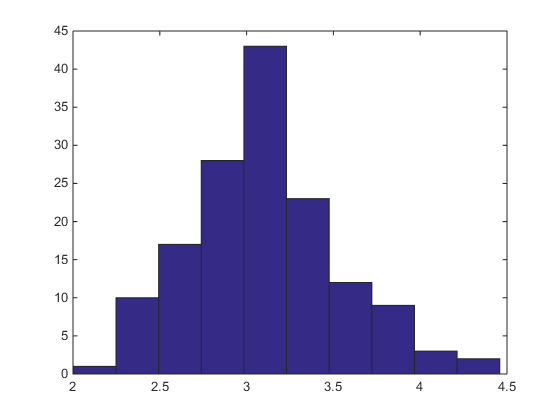
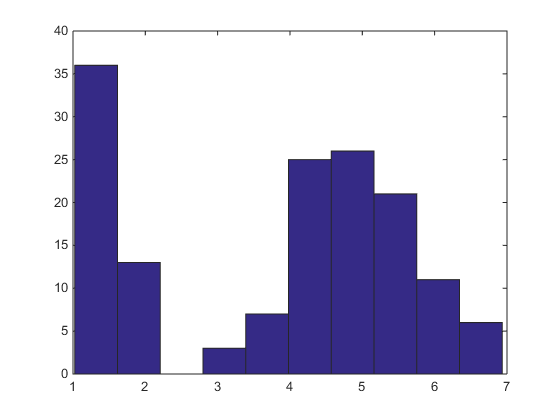
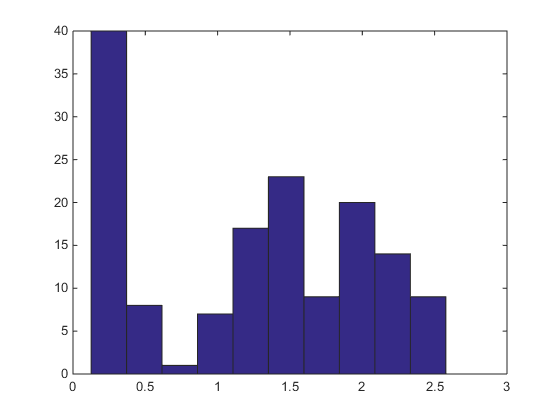
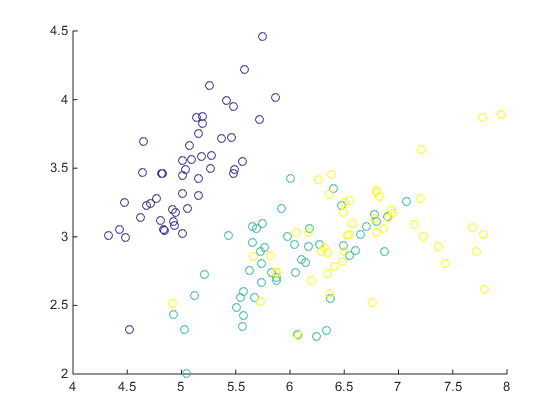
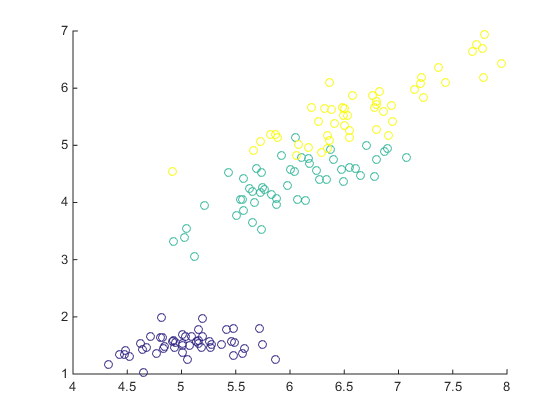
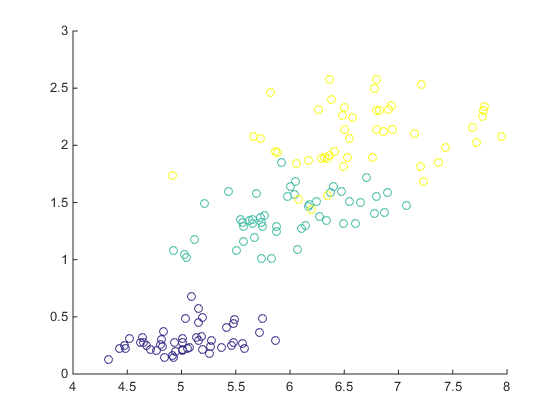
Ford Tang

46564602

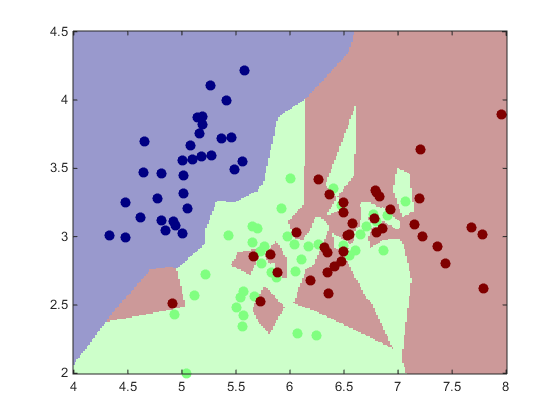
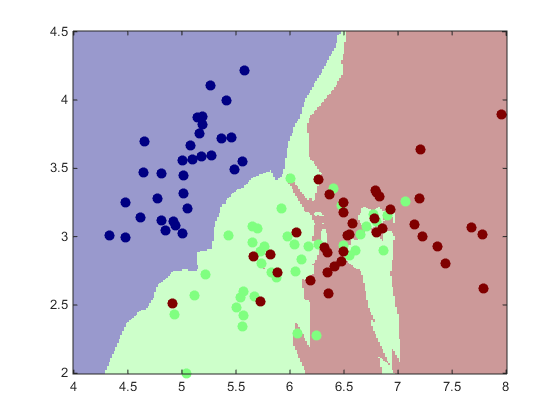
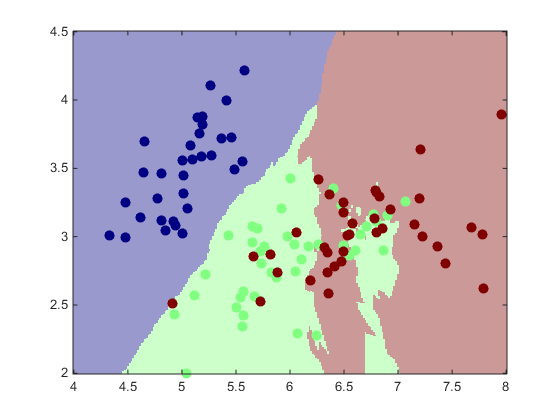
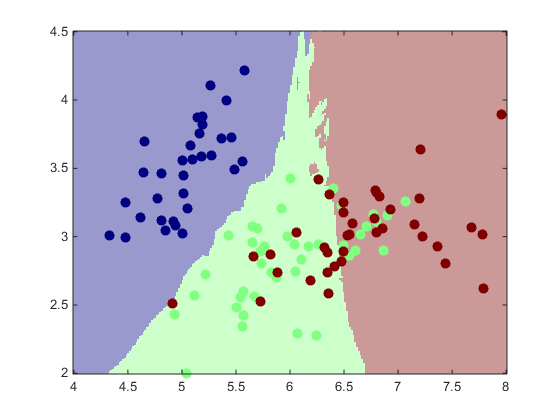
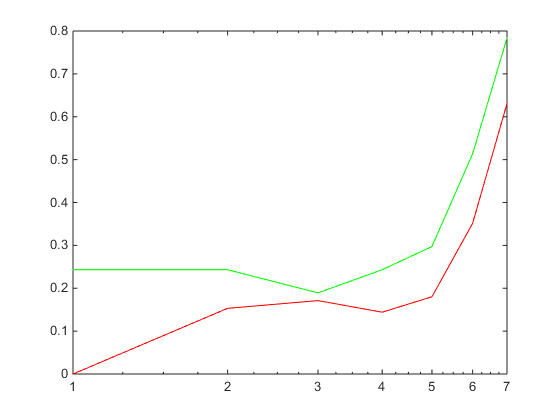
CS 178

Homework #1

Problem 1

1. size(X,2) = 4  
   size(X,1) = 148
2. hist(X(:,1)) =  
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   hist(X(:,2)) =  
     
   hist(X(:,3)) =  
     
   hist(X(:,4)) =  
   
3. mean(X) = 5.9001 3.0989 3.8196 1.2526
4. var(X) = 0.6993 0.1916 3.0976 0.5797  
   std(X) = 0.8362 0.4378 1.7600 0.7613
5. normX = X;  
   normX = bsxfun(@minus, X, mean(X));  
   normX = bsxfun(@rdivide, normX, std(normX));
6. scatter(X(:,1),X(:,2),[],y) =  
     
   scatter(X(:,1),X(:,3),[],y) =  
     
   scatter(X(:,1),X(:,4),[],y) =  
   

Problem 2

1. iris = load('data/iris.txt'); y = iris(:,end); X = iris(:,1:2);  
   [X y] = shuffleData(X,y);  
   [Xtr Xte Ytr Yte] = splitData(X,y, .75);  
   knn = knnClassify(Xtr, Ytr, 1);  
   plotClassify2D(knn, Xtr, Ytr);  
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   K = 1  
     
   knn = knnClassify(Xtr, Ytr, 5);  
   plotClassify2D(knn, Xtr, Ytr);  
   K = 5  
     
     
     
   knn = knnClassify(Xtr, Ytr, 10);  
   plotClassify2D(knn, Xtr, Ytr);  
   K = 10  
     
   knn = knnClassify(Xtr, Ytr, 50);  
   plotClassify2D(knn, Xtr, Ytr);  
   K = 50  
   
2. iris = load('data/iris.txt'); y = iris(:,end); X = iris(:,1:2);  
   [X y] = shuffleData(X,y);  
   [Xtr Xte Ytr Yte] = splitData(X,y, .75);  
   K=[1,2,5,10,50,100,200];  
   for i=1:length(K)  
   learner = knnClassify(Xtr, Ytr, K(i));  
   errTrain(i) = err(learner, Xtr, Ytr);  
   errTest(i) = err(learner, Xte, Yte);  
   end;  
   figure;  
   semilogx(errTrain, 'r');  
   hold on;  
   semilogx(errTest, 'g');  
   hold off;  
     
   K = 5 would work best.

Problem 3

1. P(X1 = 1 | y = -1) = 3/6  
   P(X2 = 1 | y = -1) = 5/6  
   P(X3 = 1 | y = -1) = 4/6  
   P(X4 = 1 | y = -1) = 5/6  
   P(X5 = 1 | y = -1) = 2/6  
     
   P(X1 = 1 | y = 1) = 3/4  
   P(X2 = 1 | y = 1) = 0/4  
   P(X3 = 1 | y = 1) = 3/4  
   P(X4 = 1 | y = 1) = 2/4  
   P(X5 = 1 | y = 1) = 1/4  
   P(y = 1) = 4/10  
   P(y = -1) = 6/10

|  |  |  |  |
| --- | --- | --- | --- |
| (X1 X2 X3 X4 X5) | P(y = -1 | X) | P(y = 1 | X) | Ŷ |
| (0 0 0 0 0) | 3/6 \* 1/6 \* 2/6 \* 1/6 \* 4/6 \* 6/10 = 0.00185 | ¼ \* 1 \* ¼ \* 2/4 \* ¾ \* 4/10 = 0.009375 | 1 |
| (1 1 0 1 0) | 3/6 \* 5/6 \* 2/6 \* 5/6 \* 4/6 \* 6/10 = 0.0463 | ¾ \* 0/4 \* ¼ \* 2/4 \* ¾ \* 4/10 = 0 | -1 |

1. P(y = 1 | (1 1 0 1 0)) = ¾ \* 0/4 \* ¼ \* 2/4 \* ¾ \* 4/10 = 0
2. With many variables, calculating dependence with Bayes classifier will be much more difficult (or impossible) and time consuming. Naïve Bayes is easier and works well enough.