# Solutions to Homework 5 (covering Statistics Lecture 7)

## Contents

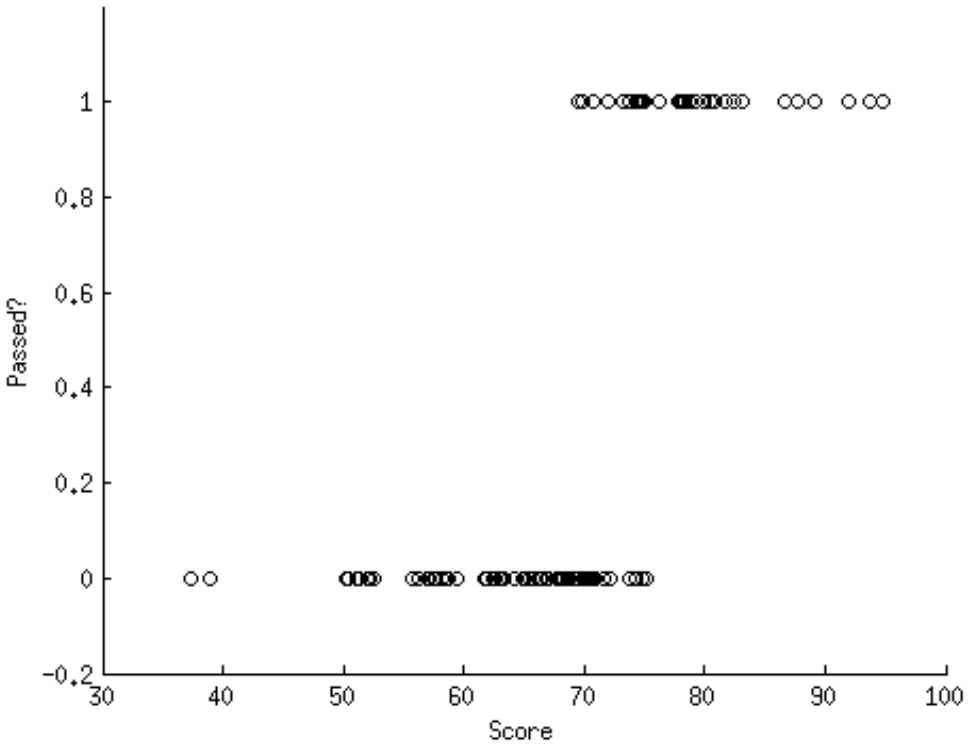


## Problem 0

```
load('Homework5.mat');
```

## Problem 1

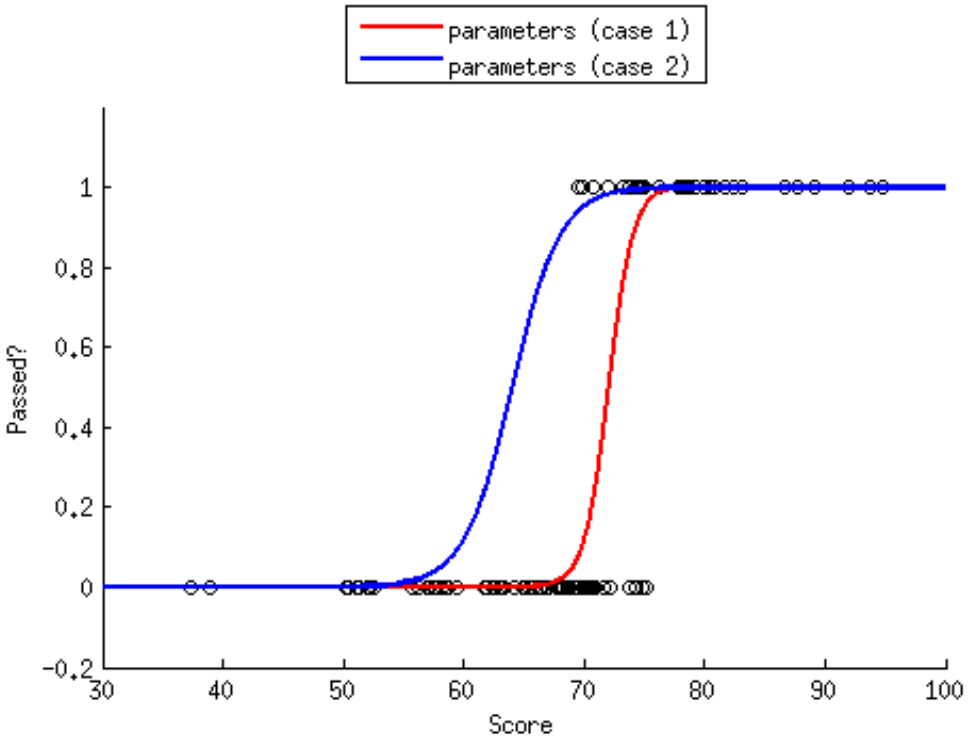
```
% visualize the data
figure; hold on;
scatter(score1,pass1,'ko');
xlabel('Score');
ylabel('Passed?');
ax = axis;
axis([ax(1:2) -.2 1.2]);
```




```
% what x-values will we evaluate models over?
ax = axis;
xx = linspace(ax(1),ax(2),100);
```

```
% define sets of parameters to try.
% each row contains a distinct set of parameters (a and b).
params = [1 -72;
          0.5 -32];
colors = {'r' 'b'};  % color to use for each set of parameters
names = {'parameters (case 1)' 'parameters (case 2)'};

% loop over parameter sets
lik =        zeros(1,size(params,1));  % likelihood of data
pctcorrect = zeros(1,size(params,1));  % percent correct achieved
h =          zeros(1,size(params,1));  % plot handle
for p=1:size(params,1)

  % define a function that takes x-values and evaluates
  % the logistic regression model at those values
  modelfun = @(xx) 1./(1+exp(-(xx*params(p,1) + params(p,2))));

  % visualize the model on the figure
  h(p) = plot(xx,modelfun(xx),[colors{p} '-'],'LineWidth',2);

  % evaluate the model for the scores that are observed
  f = modelfun(score1);

  % calculate the likelihood of the data points labeled 0
  lik0 = prod(1-f(pass1==0));

  % calculate the likelihood of the data points labeled 1
  lik1 = prod(f(pass1==1));

  % calculate the likelihood of the entire dataset
  lik(p) = lik0*lik1;

  % calculate percent correct achieved by the model
  modelfit = (f >= 0.5);
  pctcorrect(p) = sum(modelfit==pass1) / length(pass1) * 100;

  % report
  fprintf('likelihood of the data for case %d is %.g.\n',p,lik(p));

end

% add legend
legend(h,names,'Location','NorthOutside');
```

```
likelihood of the data for case 1 is 3e-11.
likelihood of the data for case 2 is 2e-44.
```




```
% which likelihood is bigger?
[mx,ix] = max(lik);

% report
fprintf('case %d has the higher likelihood.\n',ix);
fprintf('the percent correct for case %d is %.1f.\n',ix,pctcorrect(ix));
```

```
case 1 has the higher likelihood.
the percent correct for case 1 is 91.0.
```

## Problem 2

```
% calculate some variables (for code generality)
nA = size(classA,1);
nB = size(classB,1);

% construct regressor matrix
X = [classA; classB];

% construct data vector (class assignments)
y = [zeros(nA,1); ...
     ones(nB,1)];

% initialize prediction vector
pred = zeros(nA+nB,1);

% perform leave-one-out cross-validation
for p=1:length(y)
```

```
  % training and testing indices
  trainix = setdiff(1:length(y),p);
  testix = p;

  % use classify.m to perform LDA.  we train on the data points
  % in trainix and predict the class assignment for the data points
  % in testix.  we save the predicted class assignment in the
  % appropriate spot in pred.
  pred(testix) = classify(X(testix,:),X(trainix,:),y(trainix),'linear');

end

% calculate percent correct and report to window
pctcorrect = sum(pred==y) / length(y) * 100;
fprintf('class assignment successfully predicted %.1f%% of the time.\n',pctcorrect);
```

```
class assignment successfully predicted 77.1% of the time.
```

---

*Published with MATLAB® R2012b*