

# Stat 406 Spring 2010: homework 5

## 1 Mode of an Inverse Wishart distribution

(This question requires that you read the “Regularized discriminant analysis” handout.)

Show that

$$\arg \max \text{IW}(\Sigma | \mathbf{S}_0, \nu_0) = \frac{\mathbf{S}_0}{\nu_0 + D + 1} \quad (1)$$

Hint: The proof of this is very similar to the derivation of the MLE of  $\Sigma$  on p188 of the book.

## 2 MLE and MAP for multinomials

Suppose  $X \in \{1, 2\}$  and  $Y \in \{1, 2, 3\}$ . Define the joint distribution  $P(X = j, Y = k) = \theta_{j,k}$ . Consider the training data  $\mathcal{D}$  below, where row  $i$  represents  $x_i$  and  $y_i$ :

X	Y
1	1
2	2
1	3
1	1
2	2
2	3

1. Find the maximum likelihood estimates. Hint: just build the joint  $2 \times 3$  table of counts and normalize so all numbers sum to one.
2. Now suppose  $p(\theta) = \text{Dir}(\theta | \alpha_1, \dots, \alpha_6)$ , where  $\alpha_\ell = 1$  for  $\ell = 1 : 6$  (here we use  $\ell$  to represent the double index  $(j, k)$ , since we require  $\sum_\ell \theta_\ell = 1$ ). What is the MAP estimate  $\hat{\theta} = \arg \max p(\theta) p(\mathcal{D} | \theta)$ ?

## 3 Likelihood ratio for Gaussians

Consider a binary classifier where the  $K$  class conditional densities are MVN  $p(x|y = j) = \mathcal{N}(x | \mu_j, \Sigma_j)$ . By Bayes rule, we have

$$\log \frac{p(y = 1|x)}{p(y = 0|x)} = \log \frac{p(x|y = 1)}{p(x|y = 0)} + \log \frac{p(y = 1)}{p(y = 0)} \quad (2)$$

In other words, the log posterior ratio is the log likelihood ratio plus the log prior ratio. For each of the 4 cases in the table below, derive an expression for the log likelihood ratio  $\log \frac{p(x|y=1)}{p(x|y=0)}$ , simplifying as much as possible.

Form of $\Sigma_j$	Cov	Num parameters
Arbitrary	$\Sigma_j$	$Kd(d+1)/2$
Shared	$\Sigma_j = \Sigma$	$d(d+1)/2$
Shared, axis-aligned	$\Sigma_j = \Sigma$ with $\Sigma_{ij} = 0$ for $i \neq j$	$d$
Shared, spherical	$\Sigma_j = \sigma^2 I$	1

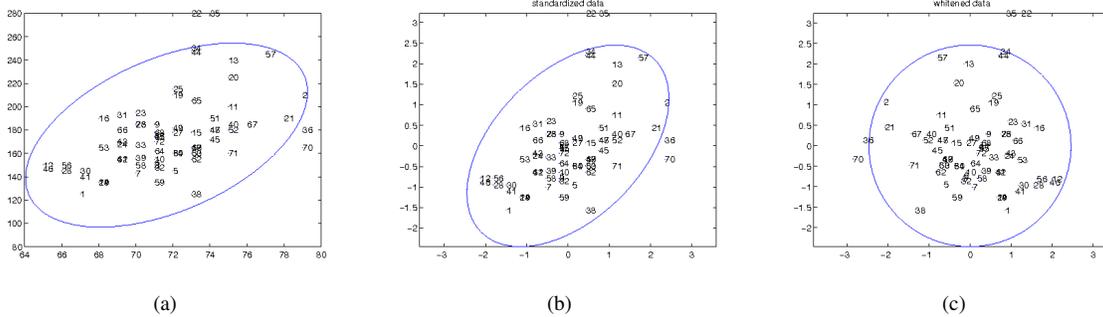


Figure 1: (a) Height/weight data for the men. (b) Standardized. (c) Whiten.

## 4 Whitening vs standardizing

(This question requires PMTKdata, available at <http://people.cs.ubc.ca/~murphyk/Data/>.)

1. Load the height/weight data using `rawdata = dlmread('heightWeightData.txt')`. The first column is the class label (1=male, 2=female), the second column is height, the third weight. Extract the height/weight data corresponding to the males. Fit a 2d Gaussian to the male data, using the empirical mean and covariance. Plot your Gaussian as an ellipse (use `gaussPlot2d`), superimposing on your scatter plot. It should look like Figure 1(a), where we have labeled each datapoint by its index. Turn in your figure and code.
2. Now standardize the data and replot. It should look like Figure 1(b). (Use `axis('equal')`.) Turn in your figure and code.
3. Now whiten the data and replot. It should look like Figure 1(c). Note that whitening rotates the data, so people move to counter-intuitive locations in the new coordinate system (see e.g., person 2, who moves from the right hand side to the left).

## 5 LDA/QDA on height/weight data

(This question requires the latest version of PMTK (5feb2010), as well as PMTKdata, available at <http://people.cs.ubc.ca/~murphyk/Data/>.)

The function `discrimAnalysisHeightWeightDemo` fits an LDA and QDA model to the height/weight data. Compute the misclassification rate of both of these models on the training set. Turn in your numbers and code.