Stat 221 Problem Set 1

Albert Young and Marco Gentili

16 September 2014

1 Question 1

1.1

We have the density of $\vec{x}$ is given by

$$f_{\vec{X}}(\vec{x}) = \frac{1}{|2\pi\Sigma|^{\frac{1}{2}}} exp\left(\frac{1}{2}(x-\vec{\mu})'\Sigma^{-1}(x-\vec{\mu}\right)$$

As in the derivation, we can represent the density of $\vec{u}$ as

$$f_{\vec{U}}(\vec{u}) = f_{\vec{X}}(g^{-1}(\vec{u}))||J||$$

The element in position (i,j) of the Jacobian is $\frac{\partial x_i}{\partial u_j}$ For diagonal elements $(i=j)$,

$$\frac{\partial x_i}{\partial u_j} = \frac{\partial}{\partial u_j}\left(\log(u_i) - \log(u_{d+1})\right) = \frac{1}{u_i} + \frac{1}{u_{d+1}}$$

For non-diagonal elements, we have

$$\frac{\partial x_i}{\partial u_j} = \frac{\partial}{\partial u_j}\left(\log(u_i) - \log(u_{d+1})\right) = \frac{1}{u_{d+1}}$$

Subtracting the each column from the first column (which preserves the determinant, we get the following.

$$det\begin{pmatrix} \frac{1}{u_1}+\frac{1}{u_{d+1}} & \frac{1}{u_{d+1}} & \dots & \frac{1}{u_{d+1}} \\ \frac{1}{u_{d+1}} & \frac{1}{u_2}+\frac{1}{u_{d+1}} & \dots & \frac{1}{u_{d+1}} \\ \dots & & & \\ \frac{1}{u_{d+1}} & \dots & \dots & \frac{1}{u_d}+\frac{1}{u_{d+1}} \end{pmatrix} = det\begin{pmatrix} \frac{1}{u_1}+\frac{1}{u_{d+1}} & \frac{1}{u_1} & \dots & \frac{1}{u_1} \\ \frac{1}{u_{d+1}} & -\frac{1}{u_2} & \dots & 0 \\ \dots & & & \\ \frac{1}{u_{d+1}} & 0 & \dots & -\frac{1}{u_d} \end{pmatrix}$$

The bottom right $(n-1)\times(n-1)$ matrix is diagonal. Multiplying each column i, where $i > 0$ by u_i/u_d and adding it to the first column, we get that the above determinant is equal to

$$det\begin{pmatrix} \frac{1}{u_1}+\frac{1}{u_{d+1}}+\frac{1}{u_1 u_{d+1}}(u_2+u_3+\ldots+u_d) & \frac{1}{u_1} & \dots & \frac{1}{u_1} \\ 0 & -\frac{1}{u_2} & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & -\frac{1}{u_d} \end{pmatrix} = det\begin{pmatrix} \frac{1}{u_1 u_{d+1}} & \frac{1}{u_1} & \dots & \frac{1}{u_1} \\ 0 & -\frac{1}{u_2} & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & -\frac{1}{u_d} \end{pmatrix}$$

This is an upper triangular matrix, so the determinant is just the product of the diagonal elements, which is equal to

$$\prod_{i=1}^{d+1} \frac{1}{u_i}$$

So then $f_{\vec{U}}(\vec{u}) = \frac{1}{|2\pi\Sigma|^{1/2} \prod_{j=1}^{d+1} u_j} \exp\left(-\frac{1}{2}\left(\log\left(\frac{\vec{u}}{u_{d+1}}\right) - \vec{\mu}\right)' \Sigma^{-1} \left(\log\left(\frac{\vec{u}}{u_{d+1}}\right) - \vec{\mu}\right)\right)$. Let us define $\frac{1}{|2\pi\Sigma|^{1/2} \prod_{j=1}^{d+1} u_j}$ to be C.

1.2

Say we have n data vectors $y_1, y_2, \dots, y_n$. Let $y_{i_{d+1}}$ be $1 - \sum_{j=1}^{d} y_j$. Then the likelihood is

$$p(y_1, y_2, \dots, y_n|\mu, \Sigma) = \prod_{i=1}^{n} C \exp\left(-\frac{1}{2}\left(\log\left(\frac{y_i}{y_{i_{d+1}}}\right) - \vec{\mu}\right)' \Sigma^{-1} \left(\log\left(\frac{y_i}{y_{i_{d+1}}}\right) - \vec{\mu}\right)\right)$$

meaning that the log likelihood is

$$\log p(y_1, y_2, \dots, y_n|\mu, \Sigma) = nC + \sum_{i=1}^{n} \left(-\frac{1}{2}\left(\log\left(\frac{y_i}{y_{i_{d+1}}}\right) - \vec{\mu}\right)' \Sigma^{-1} \left(\log\left(\frac{y_i}{y_{i_{d+1}}}\right) - \vec{\mu}\right)\right)$$

Taking the derivative of this with respect to μ and setting it equal to 0, we have

$$0 = \sum_{i=1}^{n} \Sigma^{-1}(\log\left(\frac{y_i}{y_{i_{d+1}}}\right) - \mu_{MLE})$$

Rearranging, we have that

$$\mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{y_i}{y_{i_{d+1}}}\right)$$

We can rewrite Σ as

$$\Sigma = (\alpha + \beta)\mathbb{I} - \beta\mathbb{U}$$

in which $\mathbb{I}$ is the identity matrix and $\mathbb{U}$ has ones everywhere. It holds $\mathbb{U} = \mathbb{1}\mathbb{1}'$ and then we can use Sylvester's theorem:

$$\det \Sigma = (\alpha + \beta)^d \det(\mathbb{I} - \frac{\beta}{\alpha + \beta}\mathbb{U}) = (\alpha + \beta)^d \det(1 - \frac{\beta d}{\alpha + \beta}) = (\alpha + \beta)^{d-1}\gamma$$

where $\gamma = \alpha - (d-1)\beta$. Using the Woodbury formula, we can write

$$\Sigma^{-1} = c_1\mathbb{I} + c_2\mathbb{U}$$

for some constants c_1, c_2. Noting that $\Sigma\Sigma^{-1} = \mathbb{I}$, we solve

$$c_1 = \frac{1}{\alpha + \beta}, c_2 = \frac{\beta}{(\alpha + \beta)\gamma}$$

We can then rewrite the log likelihood (looking only at terms that have α or β in them, as any others will go to 0 once we take the derivative)

$$-\frac{1}{2}\log(|\Sigma|) + (c_1 + c_2)\sum_{i=1}^{d} Var(u_i) + c_2 \sum_{1 \le i < j \le d} Cov(u_i, u_j)$$

where u_i denotes the set of all ith components of the input vectors.
Taking the derivative of the log-likelihood and setting it to zero for α and β, we get two equations and two unknowns, which we can use to solve for α and β. The algebra's kind of nasty after this point, but plugging in $\alpha = \frac{1}{n}\sum_{i=1}^{d} Var(x_i)$ and $\beta = \frac{2}{n(n-1)/2}\sum_{1\leq i<j\leq d} Cov(x_i, x_j)$, we find that these do indeed satisfy the given equations.

2

2.1

Let X denote the observed data. $X_{i,j}$ then denotes the category index of the jth feature for the ith data point. Let $\theta_{H,p,k}$ denote the probability of the kth category for feature p in the high-risk group (analogously for the low risk group). N is the number of data points and P is the number of features. Then the likelihood is

$$\prod_{i=1}^{N}\prod_{j=1}^{P}(g_{L,i}\theta_{L,j,X_{i,j}} + g_{H,i}\theta_{H,j,X_{i,j}})$$

So the log likelihood is

$$\sum_{i=1}^{N}\sum_{j=1}^{P}\log(g_{L,i}\theta_{L,j,X_{i,j}} + g_{H,i}\theta_{H,j,X_{i,j}})$$