STA 431s09 Assignment 3

Do this assignment in preparation for the quiz on Friday, Jan. 30th. Answers to Question 1 are practice for the the quiz, and are not to be handed in. For Question 2, bring both your log file and your list file to the quiz; they may (or may not) be handed in. *You should bring a calculator to the quiz.*

1. The k-dimensional multivariate normal density is

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{k}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right],$$

   where $\boldsymbol{\Sigma}$ is symmetric and positive definite.

   (a) Show that the likelihood may be simplified to

   $$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2}(2\pi)^{-nk/2} \exp -\frac{n}{2}\left[tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})\right],$$

   where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$. Show your work. You should treat the mangled answer in your lecture notes as a set of hints.

   (b) Write the log likelihood.

   (c) Any symmetric matrix is non-negative definite. Use this fact to show that for any $\boldsymbol{\Sigma}$, the log likelihood is minimized when $\boldsymbol{\mu} = \overline{\mathbf{x}}$.

   (d) You know that $\boldsymbol{\Sigma}$ positive definite implies $\boldsymbol{\Sigma}^{-1}$ positive definite. Why does this tell you that the MLE $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$ is unique?

2. In the United States, admission to university is based partly on high school marks and recommendations, and partly on applicants' performance on a standardized multiple choice test called the Scholastic Aptitude Test (SAT). The SAT has two sub-tests, Verbal and Math. A university administrator selects a random sample of 200 applicants, and obtains the Verbal SAT, the Math SAT and first-year university Grade Point Average (GPA) for each student. She wants an equation for predicting GPA from Verbal SAT and Math SAT. That is, predicted GPA will be a function of two variables, Verbal SAT and Math SAT. The raw data are available in the file sat.data; there is a link on the course website in case the one from this document does not work.

Write a SAS program and produce output that will allow you to answer questions such as the following from your printout. You should bring a calculator to the quiz.

(a) What proportion of the sample variation in GPA is explained by the two components of the SAT test? The answer to this question is a single number between zero and one.

(b) Write down the prediction equation based on material in your list file. Denote predicted GPA by $\widehat{Y}$, Verbal SAT by X_1 and Math SAT by X_2. All other components of the formula should be numbers.

(c) Give predicted GPA for a new student with a verbal SAT of 600 and a math SAT of 650. The answer is a single number. To answer questions like this on the quiz, it will help to have a calculator.

Please bring hard copy of both your log file and your list file to the quiz. Please note that

- If your log file contains errors or warnings, you must track them down and eliminate them.

- The log file contains a listing of your program, but it is not just a listing of your program. You are *not* being asked to bring a copy of your program file (sat.sas or whatever) to the quiz.

- The log file and the list file you bring to the quiz *must* come from the same run. This allows us to trace any errors you might have, and you will lose substantial marks if you ignore this request.

- To do this assignment (and all SAS assignments in this class), you must use your computer account on tuzo, and *only* your own account. As stated on the course outline, the first time you violate this rule, the mark for both students on the computer part of the quiz will be zero, even if there is no evidence of copying. The second time, Computing Services will be asked to lock the computer account.

- SAS can't easily deal with the variable names above the columns, so you should delete the first three rows of the data file before using it.