Applied Statistics Exam
May 2008
Time: 12:00pm–3:30pm

This exam consists of 2 parts. You must answer 5 questions total and must answer at least 2 questions from each part. Make sure to clearly indicate which problems you are attempting. Some formulas and tables are given at the end of this exam.

PART A:

A1. Suppose that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{y}$ is a N-dimensional column vector of outputs, $\boldsymbol{X}$ is a $N \times p$ design matrix of fixed inputs where $rank(\boldsymbol{X}) = p$ and $p < N$, $\boldsymbol{\beta}$ is a p-dimensional column vector of coefficients, and $\boldsymbol{\epsilon}$ follows a Normal distribution with mean vector $\mathbf{0}_N$ and covariance matrix $\sigma^2 \boldsymbol{I}_N$.
(a) Let $Q(\boldsymbol{b}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}\|^2$. Compute $\dfrac{\partial Q}{\partial \boldsymbol{b}}$.
(b) Derive the solution to the score equation $\dfrac{\partial Q}{\partial \boldsymbol{b}} = \mathbf{0}_p$ and denote it by $\hat{\boldsymbol{\beta}}$.
(c) Find the distribution of $\hat{\boldsymbol{\beta}}$.
(d) Find the distribution of the residual vector $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$.
(e) Show that $\hat{\boldsymbol{\beta}}$ and $RSS = \boldsymbol{r}^\top \boldsymbol{r}$ are independent.

A2. Consider the simple linear regression model in which $\boldsymbol{y} \sim \text{Normal}_N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_N)$ where $\boldsymbol{X} = [\boldsymbol{J} \vdots \boldsymbol{x}]$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$; here β_0 is the unknown fixed intercept parameter, β_1 is the unknown fixed slope parameter, σ^2 is the unknown fixed variance, $\boldsymbol{x}$ is a known non-random N-dimensional column vector, and $\boldsymbol{J}$ is a N-dimensional column vector of ones.

Assume that we have the following summary statistics:

$$N = \boldsymbol{J}^\top \boldsymbol{J} = 10,\ \boldsymbol{J}^\top \boldsymbol{x} = 20,\ \boldsymbol{x}^\top \boldsymbol{x} = 100,\ \boldsymbol{J}^\top \boldsymbol{y} = 0, \boldsymbol{x}^\top \boldsymbol{y} = -10,\ \boldsymbol{y}^\top \boldsymbol{y} = 46.$$

(a) Compute the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2. Denote them as $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, respectively.
(b) What is the distribution of $\dfrac{\|\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2}{2s^2}$, where $s^2 = \frac{N}{N-2}\hat{\sigma}^2$?
(c) Find a 95% confidence ellipse for $\boldsymbol{\beta}$ and find values A, B, and C so that the confidence ellipse can be expressed in the form

$$A(\hat{\beta}_0 - \beta_0)^2 + B(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + C(\hat{\beta}_1 - \beta_1)^2 \leq 1.$$

A3. Consider the regression model

$$y_i \sim \begin{cases} \text{Normal}(\mu_1 + \beta x_i, \sigma^2) & \text{for } i = 1, \ldots, n_1 \\ \text{Normal}(\mu_2 + \beta x_i, \sigma^2) & \text{for } i = n_1 + 1, \ldots, n_1 + n_2 \end{cases}$$

where μ_1, μ_2, and β are unknown fixed regression parameters, σ^2 is the unknown fixed variance, n_1 and n_2 are the known number of observations in groups 1 and 2, respectively, and $x_i, i = 1, \ldots, n_1 + n_2$, are known and non-random.

Suppose that $n_1 = 3$, $n_2 = 3$, and the data are given in the following table:

i	1	2	3	4	5	6
x_i	-1	0	1	-1	-1	2
y_i	1	3	-3	0	4	2

(a) Compute the maximum likelihood estimates of μ_1, μ_2, β, and σ^2.
(b) Compute an appropriate test statistic for testing $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$, and determine whether H_0 should be rejected when tested at level $\alpha = .05$.

A4. Consider the design matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ 1 & x_2 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^d \end{bmatrix}$$

used in a polynomial regression model. Assume that $N > d + 1$ and $\boldsymbol{X}$ is full rank.
(a) Show that the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is symmetric and idempotent and that $\boldsymbol{HX} = \boldsymbol{X}$.
(b) Let h_{ij} denote the element in the ith row and jth column of $\boldsymbol{H}$. Using the result of part (a), show that $\sum_{j=1}^{N} h_{ij} = 1$ for $i = 1, \ldots, N$.
(c) Let $\chi_0 = [1, x_0, x_0^2, \cdots, x_0^d]$ be a $(d+1)$-dimensional row vector used for modeling a new input $x_0 \neq 0$, and let $b(\chi_0) = \chi_0(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top = [b_1, \cdots, b_N]$. Show that $\sum_{i=1}^{N} b_i x_i^k = x_0^k$ for $k = 0, 1, \ldots, d$.

(d) Use the result in (c) to show that $\sum_{i=1}^{N} b_i (x_i - x_0)^k = 0$ for $k = 0, 1, \ldots, d$.

PART B:

B1. Suppose each of K classes has an associated target t_k, which is a vector of all zeros, except a one in the kth position. Show that classifying to the largest element of $\hat{y}$ amounts to choosing the closest target, $\min_k \|t_k - \hat{y}\|$, if the elements of $\hat{y}$ sum to one.

B2. Consider a logistic regression model with no intercept

$$\ln \frac{\mathrm{P}(G = 1|X = x)}{\mathrm{P}(G = 0|X = x)} = \beta x$$

with a real-valued input variable X used to classify the Bernoulli output variable G, and suppose that there are N independent observations

$$(x_1, g_1), \ldots, (x_N, g_N)$$

that can be used to fit the model.

(a) Derive an expression for the log-likelihood function

$$\ell(\beta) = \sum_{i=1}^{N} \ln \mathrm{P}(G = g_i | X = x_i).$$

(b) Show that

$$\frac{d\ell}{d\beta} = \sum_{i=1}^{N} x_i \left(g_i - \frac{e^{\beta x_i}}{1 + e^{\beta x_i}} \right).$$

(c) Consider an experiment which attempts to use a single input variable (temperature) to model the probability that a response is a success; assume it is known that 50% of the responses are expected to be successes when the input variable is set at 0°C so that the no intercept model is appropriate. Suppose that repeated trials are performed at two different temperatures −1°C and 1°C, and it is seen that there are n_1 failures and n_2 successes at −1°C while there are n_3 failures and n_4 successes at 1°C; that is, $N = n_1 + n_2 + n_3 + n_4$ observations for this experiment are

$$\underbrace{(-1,0), \ldots, (-1,0)}_{n_1}, \underbrace{(-1,1), \ldots, (-1,1)}_{n_2}, \underbrace{(1,0), \ldots, (1,0)}_{n_3}, \underbrace{(1,1), \ldots, (1,1)}_{n_4}.$$

Compute the maximum likelihood estimate of β for a logistic regression model with no intercept based on this data.

B3. Suppose $y_i = f(x_i) + \epsilon_i$ for $i = 1, \ldots, N$ where the x_i's are fixed and distinct and $\epsilon_1, \ldots, \epsilon_N$ are i.i.d. with $\mathrm{E}[\epsilon_i] = 0$ and $var[\epsilon_i] = \sigma^2$. Consider the kernel-weighted average

$$\hat{f}_\lambda(x) = \frac{\sum_{i=1}^N K_\lambda(x, x_i) y_i}{\sum_{i=1}^N K_\lambda(x, x_i)}$$

with the Gaussian kernel

$$K_\lambda(x, x_i) = \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{1}{2}\left(\frac{x - x_i}{\lambda}\right)^2}.$$

Suppose we want to estimate the response at the first observation time x_1.
(a) Compute $\lim_{\lambda \to 0^+} \hat{f}_\lambda(x_1)$ and $\lim_{\lambda \to \infty} \hat{f}_\lambda(x_1)$.
(b) Find the expected value of $\hat{f}_\lambda(x_1)$, and compute $\lim_{\lambda \to 0^+} \mathrm{E}[\hat{f}_\lambda(x_1)]$ and $\lim_{\lambda \to \infty} \mathrm{E}[\hat{f}_\lambda(x_1)]$.
(c) Find the variance of $\hat{f}_\lambda(x_1)$, and compute $\lim_{\lambda \to 0^+} var[\hat{f}_\lambda(x_1)]$ and $\lim_{\lambda \to \infty} var[\hat{f}_\lambda(x_1)]$.

B4. Consider the following univariate data: 0, 3, 7, 8, 13. Use the squared Euclidean distance $d(x, y) = (x - y)^2$ for all parts.
(a) Apply the K-means algorithm to obtain $K = 2$ clusters by beginning with cluster A centered at 0 and cluster B centered at 3. Compute the within cluster scatter for the final configuration.
(b) Draw the dendogram for the single linkage agglomerative clustering strategy. How would this strategy partition the data into two groups?
(c) Draw the dendogram for the complete linkage agglomerative clustering strategy. How would this strategy partition the data into two groups?

FORMULAS:

Suppose $\boldsymbol{y} \sim \text{Normal}_N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_N)$, $\boldsymbol{X}$ is a $N \times p$ full rank matrix, $N > p$, and $\boldsymbol{X}^\top \boldsymbol{X}$ is invertible. Let $\hat{\boldsymbol{\beta}}$ be the MLE of $\boldsymbol{\beta}$ and let $\hat{\boldsymbol{\beta}}_0$ be the restricted MLE of $\boldsymbol{\beta}$ satisfying $\boldsymbol{K}^\top \hat{\boldsymbol{\beta}}_0 = \boldsymbol{m}$. If $\boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$, then

$$\begin{aligned} F &= \frac{(\text{reduced SS} - \text{full SS})/q}{\text{full SS}/(N-p)} \\ &= \frac{\|\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0)\|^2/q}{\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2/(N-p)} \\ &= \frac{(\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})^\top (\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})/q}{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/(N-p)} \sim f_{q,N-p} \end{aligned}$$

where reduced SS$= \|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0\|^2$ and full SS$= \|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2$.

TABLES:

1. The 100αth percentage point of the central t-distribution with *df* degrees of freedom.

2. The 100αth percentage point of the central χ^2-distribution with *df* degrees of freedom.

3. Upper α probability points of the central F-distribution with n_1 d.f. in the numerator and n_2 d.f. in the denominator.