LINEAR REGRESSION MODELS W4315

Midterm Examination QUESTIONS
November 19, 2010

Instructor: Frank Wood (10:35-11:50)

- Percentile table for t distribution.

Percentile	0.9	0.95	0.975	0.9875
df=3	1.64	2.35	3.18	4.18
df=4	1.53	2.13	2.78	3.50
df=5	1.48	2.02	2.57	3.16

- Percentile table for F distribution.

Percentile	0.9	0.95	0.975	0.9875
F(1,3)	5.54	10.13	17.44	29.07
F(1,4)	4.54	7.71	12.22	18.62
F(1,5)	4.06	6.61	10.01	14.52

1. (40 points) The data below give weight X (kg) and height Y (cm) of 5 teenagers.

X (kg)	Y (cm)
50	160
60	160
70	170
80	170
90	180

Consider a simple linear regression model specification with errors ϵ_i i.i.d. normally distributed with mean 0 and variance σ^2. The least square estimators are given by

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$
$$b_0 = \bar{Y} - b_1 \bar{X}$$

1. (10 pts) What's the sampling distribution of b_1?

2. (10 pts) Find a 95% confidence interval for β_1.

3. (10 pts) Test one-sided hypothesis $H_0 : \beta_1 \leq 0$ versus $H_a : \beta_1 > 0$ (level of significance $\alpha = 0.05$)

4. (10 pts) Produce an ANOVA table and perform an F-test for $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. (level of significance: $\alpha = 0.05$).

Source of Variation	SS	df	MS
Regression			
Error			
Total			

2. (30 points) Consider the classical regression setup

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

where $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$. We know that the least square estimator for the parameter β minimizes the residual sum of squares, which in matrix terms can be written $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$. The value of β which minimizes this expression has the following analytic form

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{1}$$

Suppose, though, that $\mathbf{X}'\mathbf{X}$ is not invertible. In this case, this estimator can't be used. To get around this problem we define a penalized residual sum of squares (this is called "ridge regression")

$$(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta'\beta. \tag{2}$$

(a) (20 pts) Derive the ridge regression estimate $\hat{\beta}^{ridge}$ in matrix form (show your work), this is equivalent to finding the β which minimizes (2).

(b) (10 pts) In no more than 2 sentences, explain why the resulting estimator will work even when $\mathbf{X}'\mathbf{X}$ is singular.

3. (40 points) You, a statistical analysis consultant, are asked to analyze a regression problem. Food scientists invent a new secret ingredient to increase the fluffiness of pancakes (fluffiness is a made up concept but for our purposes will be defined as measurable scalar

quantity). They know that the concentration of the new ingredient affects the fluffiness, and they want to figure out how large the effect is. We assume that a linear regression model is appropriate and the error comes from measurement of fluffiness. Two factories experimentally produce the pancakes with varying concentrations of the new ingredient. It could have been a simple linear regression problem, but the difficulty here is that the variances of fluffiness measurement of the pancakes produced by the two factories are DIFFERENT (that is why you are called). In other words, the fluffiness measurement errors were iid only for each factory individually. Denote fluffiness as $y = [y_1^1, y_1^2, \ldots, y_1^{n_1}, y_2^1, y_2^2, \ldots, y_2^{n_2}]$ and concentration level as $x = [x_1^1, x_1^2, \ldots, x_1^{n_1}, x_2^1, x_2^2, \ldots, x_2^{n_2}]$, where subscripts denote factories.

Hint : here because the i.i.d. assumption is no longer valid, instead of only one parameter σ^2, we need to use two parameters σ_1^2 and σ_2^2. You may find it easier, after expressing the problem in matrix form, to simplify the likelihoods by expressing them in scalar form before seeking the ML estimates for some of the variables.

(a) (15 pts) Assume (for now) that the variances σ_1^2 and σ_2^2 are *known*. Using matrix notation derive a maximum likelihood estimator for the vector of regression coefficients $\vec{\beta} = [\beta_0, \beta_1]^T$, i.e. $\hat{\beta} = ?$.

(b) (15 pts) Now assume that σ_1^2 and σ_2^2 are unknown. Use maximum likelihood techniques to derive estimators for both (i.e. $\hat{\sigma}_1^2 = ?$ and $\hat{\sigma}_2^2 = ?$. You can plug $\hat{\beta}$ in for β in the final step.

(c) (10 pts) Since the answer to (b) above forms a coupled system of equations, in two or fewer sentences describe a way to use the coupled equations to estimate all unknown parameters.