

Statistics 550 Homework assignment 5

1. Exercise 13.7 (Hint: Use the result that for two nonsingular $k \times k$ matrices A and B , the matrix AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$)
2. For the linked baseball data set, fit a multiple regression model to predict salary in 1987 (salary87) from at bats, hits, homeruns, runs scored, runs batted in, and walks in 1986 (ab86 h86 r86 hr86 r86 rbi86 w86) and years of major league service (years). Calculate variance-inflation factors and perform a principal components analysis of the covariates to assess collinearity.
3. For the baseball data above, use stepwise selection and best subsets regression methods (with MSE, AIC, and BIC) to investigate models for predicting salary87. Use the same covariates as in the problem above. Which model(s) is(are) best?
4. Use a ridge regression estimator for the baseball salary prediction model from problem 2. Pick a value for the ridge constant d by looking at the ridge trace figure - how does your estimate of β compare to the estimate from problem 2?
5. For the second data set on nutrient uptake, fit the Michaelis-Menten kinetic model to predict nutrient uptake (y) from substrate concentration (x),

$$y = \frac{V_{\max}}{k + x},$$

where V_{\max} is maximum velocity and k is the transport constant. Using a plot of the data, obtain starting values for V_{\max} and k then use Proc NLIN to obtain parameter estimates and asymptotic standard errors.

6. Exercise 14.7.
7. For the linked baseball data set, fit a logistic regression model to predict if a player is a shortstop using the set of covariates from Problem 2 above. What is the predicted probability of a player being a shortstop if he had 500 at bats, 125 hits, 3 home runs, 87 runs scored, 32 runs batted in, and 22 walks in 1986 and had 5 years of major league service?
8. Exercise 15.5.