Homework 2: STA 721 Fall19

Your Name

September 9, 2019

1. Consider the linear model $\mathbf{Y} \sim \ N(\mu\,\sigma^2\mathbf{I}_n)$ with $\boldsymbol{\mu} = \mathbf{1}_n\beta_0 + \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ a full rank matrix with (column) rank p, where $\mathbf{X}$ is linearly independent of the vector $\mathbf{1}_n$.

   (a) Show that the projection, $\mathbf{P}$, on the column space spanned by the vector $\mathbf{1}_n$ of length n and $\mathbf{X}$ may be written as
   $$\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_{\mathbf{X}-\mathbf{1}_n\bar{\mathbf{x}}^T}$$
   where $\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}$. Show that diagonal elements are
   $$h_{ii} = \frac{1}{n} + (\mathbf{x}_i - \bar{\mathbf{x}})^T\left((\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}^T)\right)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$$
   (recall all vectors are column vectors). The h_{ii} are known as the leverage values. *Hint:* show that the mean function can be re-written as
   $$\begin{aligned}\boldsymbol{\mu} =& \mathbf{1}_n\beta_0 + \mathbf{X}\boldsymbol{\beta} \\ =& \mathbf{1}_n\beta_0 + \mathbf{P}_1\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}\boldsymbol{\beta} \\ =& \mathbf{1}_n\beta_0 + \mathbf{1}_n\bar{x}^T\boldsymbol{\beta} + (X - \mathbf{1}_n\bar{x}^T)\boldsymbol{\beta} \\ =& \mathbf{1}_n\alpha_0 + \mathbf{X}_c\boldsymbol{\beta}\end{aligned}$$
   where $\mathbf{X}_c = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}$ is the centered matrix of predictors and
   $$\alpha = \mathbf{1}_n\beta_0 + \mathbf{1}_n\bar{\mathbf{x}}^T\boldsymbol{\beta}$$
   is the intercept in the model with the centered parameters. If the mean is in the column space of $C(\mathbf{1}_n, \mathbf{X})$ show that it is in the column space $C(\mathbf{1}_n, \mathbf{X}_c)$. Last show that $\mathbf{P}_1 + \mathbf{P}_{\mathbf{X}_c}$ is an orthogonal projection on the column space $C(\mathbf{1}_n, \mathbf{X})$.

   (b) Find the sampling distribution of $\hat{\mu}_i$ (the mean of Y_i at $\mathbf{x}_i^T$) and express the variance as a function of h_{ii}. Provide an expression for a 95% confidence interval. For what values of $\mathbf{x}$ will the interval be the narrowest? Explain.

   (c) Given σ^2, find the distribution of e_i as a function of h_{ii}. Explain (rigorously) why e_i unconditional on σ^2 does not have a student t distribution with $n - p - 1$ degrees of freedom.

2. Now consider predicting Y_* at a new point $\mathbf{x}_*^T$ where $Y_* \sim N(1\beta_0 + \mathbf{x}_*^T\boldsymbol{\beta}, \sigma^2)$.

   (a) Find the distribution of the predicted residual $e_* = Y_* - 1\hat{\beta}_0 - \mathbf{x}_*^T\hat{\boldsymbol{\beta}}$ (given $\beta_0, \boldsymbol{\beta}$ and σ^2). Both Y and Y_* are random variables here. *Hint: consider using the centered parameterization*

   (b) Show that the standardized predicted residual (center so that the mean is 0 and and scale (sd) is 1 with σ^2 replaced by the usual unbiased estimate $\hat{\sigma}^2 = \mathbf{Y}^T(\mathbf{I}_n\mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_c})\mathbf{Y}/(n-p-1)$ has a student t distribution. What are the degrees of freedom? (Explain)

   (c) Use the standardized predicted residual to construct a 95% Confidence interval (also called a prediction interval) for Y_*.

3. Refer to the Prostate data from `library(lasso2) data(Prostate, package="lasso2")` for this proplem (see R code from Lecture.)

   (a) Fit the regression model with response `lcavol`, and variables `svi` and `lpsa` as predictors. Construct 95% confidence intervals for each coefficient and provide a meaningful interpretations for changes in the median cancer volume (not log cancer volume) include any units etc in your interpretation. Note "a 1 unit" change may or may not be meaningful for interpretation so adjust as needed.

   (b) Plot the cancer volume versus `PSA` using a log scale on both axes. Add the fitted regression function for `svi = 1` and `svi = 0`, with lines representing the (pointwise) 95% confidence intervals (CI) for each. Use a different color and line type for the fitted function and the confidence intervals. *Hint: see the* `predict` *function in R to obtain the confidence intervals*

   (c) Add to the plot 95% prediction intervals from `predict` using a different line type and color from the CI. Add a legend to your plot and an informative caption.

   (d) Explain why are the prediction intervals wider than the confidence intervals. Where will the prediction intervals be the narrowest? Will the width of the prediction intervals ever go to 0, with an increasing sample size? Explain.

4. Consider the model
$$\mathbf{Y} \mid \mathbf{X} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$
and estimation of $\boldsymbol{\beta}$ using quadratic loss $(\boldsymbol{\beta} - \mathbf{a})^T(\boldsymbol{\beta} - \mathbf{a})$ for some estimator $\mathbf{a}$. Find the expected quadratic loss if we use the MLE $\hat{\boldsymbol{\beta}}$ for $\mathbf{a}$ conditional on $\mathbf{X}$. Simplify the expression as a function of the eigenvalues of $\mathbf{X}^T\mathbf{X}$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?

5. Consider estimation of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ at the observed data points $\mathbf{X}$. Find the expected quadratic loss $E[(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})]$ conditional on $\mathbf{X}$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?

6. Consider predicting a new $\mathbf{Y}_*$ at the observed data points $\mathbf{X}$ where $\mathbf{Y}_*$ is independent of $\mathbf{Y}$. Find the expected quadratic loss for $E[(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})]$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0?

7. Consider predicting $\mathbf{Y}_*$'s at new points $\mathbf{X}_*$ with $E[\mathbf{X}_*^T\mathbf{X}_*] = \mathbf{I}_p$. Find the expected quadratic loss $E[(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})]$ conditional on $\mathbf{X}$ and $\mathbf{X}_*$ and then unconditional on $\mathbf{X}_*$ (but still conditional on $\mathbf{X}$. What happens as the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$ goes to 0? (If $E[\mathbf{X}_*^T\mathbf{X}_*] = \Sigma_{\mathbf{X}} > 0$ does that change the result)

8. Briefly comment on the difference in estimation and prediction at observed data versus new data as $\mathbf{X}$ becomes non-full rank. Which is the most stable? Which is the least?