BST 140.751
Problem Set 2
Due: September 21, 2017

1 Least squares

1. Suppose that the design matrix is of the form $\mathbf{J}_A \otimes \mathbf{I}$ where $\mathbf{J}_A$ is a vector of length A and $\mathbf{I}$ is a $B \times B$ identity matrix. Let $\mathbf{y}$ be a $AB \times 1$ length vector. Find the least squares estimates associated with this design matrix.

2. Let $Y_{ij} = \alpha_0 + \beta_j + \epsilon_{ij}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.
   A. Write out the design matrix for the associated linear model.
   B. Show what the estimates are under the following constraints:
      i. $\alpha_0 = 0$
      ii. $\beta_1 = 0$
      iii. $\beta_J = 0$
      iv. $\sum_{j=1}^{J} \beta_j = 0$

3. Let $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ where $\mathbf{X}_1$ is $n \times p_1$ and $\mathbf{X}_2$ is $n \times p_2$. Consider minimizing $||\mathbf{y} - \mathbf{X}\beta||$ where $\beta = (\beta_1'\ \beta_2')'$. Show that when $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal, the estimate of β_1 does not change when $\mathbf{X}_2$ varies.

4. Consider the model $\mathbf{y} = \mathbf{X}\beta + \epsilon$. Let $\mathbf{F}$ be an invertible $p \times p$ matrix and $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{F}$.
   A. Consider another model $\mathbf{y} = \tilde{\mathbf{X}}\tilde{\beta} + \epsilon$. Argue that the models are equivalent.
   B. Show that the least squares estimate of $\tilde{\beta}$ from the second model is $\mathbf{F}^{-1}\hat{\beta}$ where $\hat{\beta}$ is the least squares estimate from the first model.
   C. Suppose that you have a linear regression equation where one of the regressors is temperature. Use the results above to relate the beta coefficients if the regressor is input as Celsius or Fahrenheit.

5. Assume that $\mathbf{P}_i$ is a projection matrix ($i = 1, 2$) and $\mathbf{P}_1 - \mathbf{P}_2$ is positive semidefinite. Show that:
   A. $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_2$ (Hint: first show that $\mathbf{P}_1\mathbf{x} = \mathbf{0}$ implies $\mathbf{P}_2\mathbf{x} = \mathbf{0}$).
   B. $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix.

2 Computing and analysis

1. Extend the `R` function `mylm()` you created in HW1 to do the following:

   A. Check to make sure that X and Y have the right format (matrix and vector, respectively), have the right dimensions, have no missing values, NA or Inf, are numeric and that X is full rank.

   B. Return a list with the following information:

      i. beta, the vector of least squares estimates,
      ii. fitted, the vector of fitted values,
      iii. residuals, the vector of residuals.
      iv. R2, the value of R-squared.
      v. hatdiag, the vector of the diagonals of the hat matrix.

2. Download the file `HW2ex.txt` from the Courseplus site. Let the first column be the response variable, and the remaining four columns the explanatory variables. Use your function to obtain the fitted values and residuals. Hand in a plot of predicted values against the residuals.

Advanced Methods Homework 2

Bohao Tang

September 15, 2017

1 Least Squares

1. $\mathbf{J}_A \otimes \mathbf{I}$ is like $\begin{pmatrix}\mathbf{I} & \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I}\end{pmatrix}^\top$. Suppose the model is $\mathbf{y} = (\mathbf{J}_A \otimes \mathbf{I})\,\beta + \boldsymbol{\epsilon}$. Then we have the estimation $\hat{\beta}$ is $((\mathbf{J}_A \otimes \mathbf{I})^\top \mathbf{J}_A \otimes \mathbf{I})^{-1}(\mathbf{J}_A \otimes \mathbf{I})^\top\ \mathbf{y}$. And it is:

$$\begin{aligned}
\hat{\beta} &= \left(\begin{pmatrix}\mathbf{I} & \mathbf{I} & \cdots & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{I}\\ \mathbf{I}\\ \vdots\\ \mathbf{I}\end{pmatrix}\right)^{-1}\begin{pmatrix}\mathbf{I} & \mathbf{I} & \cdots & \mathbf{I}\end{pmatrix}\mathbf{y} && (1)\\
&= \frac{1}{A}\mathbf{I}\sum_{j=1}^{A}\mathbf{y}_i = \sum_{j=1}^{A}\mathbf{y}_i/A && (2)
\end{aligned}$$

Where $\mathbf{y}_i$ is vectors of length B and $\mathbf{y} = \begin{pmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_A\end{pmatrix}^\top$

2. A.

$$\begin{pmatrix}Y_{11}\\ Y_{12}\\ Y_{13}\\ \vdots\\ Y_{1J}\\ Y_{21}\\ Y_{22}\\ \vdots\\ Y_{2J}\\ \vdots\\ Y_{IJ}\end{pmatrix} = \begin{pmatrix}1 & 1 & & & & \\ 1 & & 1 & & & \\ 1 & & & 1 & & \\ \vdots & & & & \ddots & \\ 1 & & & & & 1\\ 1 & 1 & & & & \\ 1 & & 1 & & & \\ 1 & & & 1 & & \\ \vdots & & & & \ddots & \\ 1 & & & & & 1\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & 1 & & & & \\ 1 & & 1 & & & \\ 1 & & & 1 & & \\ \vdots & & & & \ddots & \\ 1 & & & & & 1\end{pmatrix}\begin{pmatrix}\alpha_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_J\end{pmatrix} + \boldsymbol{\epsilon} = \begin{bmatrix}\mathbf{J}_{IJ} & \mathbf{J}_I \otimes \mathbf{I}_J\end{bmatrix}\begin{pmatrix}\alpha_0\\ \boldsymbol{\beta}\end{pmatrix} + \boldsymbol{\epsilon}$$

So the design matrix is $\begin{bmatrix}\mathbf{J}_{IJ} & \mathbf{J}_I \otimes \mathbf{I}_J\end{bmatrix}$, where $\mathbf{I}_J$ is the $J \times J$ identity matrix.

B. i. Then the design matrix is $\mathbf{J}_I \otimes \mathbf{I}_J$, denote $\mathbf{Y}_i = \begin{bmatrix}Y_{i1} & Y_{i1} & \cdots & Y_{iJ}\end{bmatrix}^\top$. According to question one, we have that $\hat{\beta} = \sum_{j=1}^{I}\mathbf{y}_i/I$.

ii. Then the design matrix is $D = \begin{bmatrix}\mathbf{J}_{IJ} & \mathbf{J}_I \otimes \mathbf{L}\end{bmatrix}$, where $\mathbf{L} = \begin{pmatrix}\mathbf{0}\\ \mathbf{I}_{J-1}\end{pmatrix}$. So the estimation is (denote $\mathbf{Y}_{-1,i} = (Y_{i2}, Y_{i3}, \cdots, Y_{iJ})^\top$):

$$\begin{aligned}
\widehat{(\alpha_0, \beta_2, \cdots, \beta_J)}^\top &= (D^\top D)^{-1} D^\top \mathbf{Y} &(3)\\
&= \frac{1}{I}\begin{pmatrix} J & \mathbf{J}_{J-1}^\top \\ \mathbf{J}_{J-1} & \mathbf{I}_{J-1}\end{pmatrix}^{-1} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I \mathbf{Y}_{-1,i}\end{pmatrix} &(4)\\
&= \frac{1}{I}\begin{pmatrix} 1 & -\mathbf{J}_{J-1}^\top \\ -\mathbf{J}_{J-1} & \mathbf{I}_{J-1} + \mathbf{J}_{J-1}\mathbf{J}_{J-1}^\top\end{pmatrix} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I \mathbf{Y}_{-1,i}\end{pmatrix} &(5)\\
&= \begin{pmatrix}\sum_{i=1}^I Y_{i1}/I \\ (\sum_{i=1}^I \mathbf{Y}_{-1,i} - \sum_{i=1}^I Y_{i1}\mathbf{J}_{J-1})/I\end{pmatrix} &(6)
\end{aligned}$$

iii. Then the design matrix is $D = \begin{bmatrix}\mathbf{J}_{IJ} & \mathbf{J}_I \otimes \mathbf{L}_2\end{bmatrix}$, where $\mathbf{L}_2 = \begin{pmatrix}\mathbf{I}_{J-1}\\ \mathbf{0}\end{pmatrix}$. So the estimation is (denote $\mathbf{Y}_{-J,i} = (Y_{i1}, Y_{i2}, \cdots, Y_{i,J-1})^\top$):

$$\begin{aligned}
\widehat{(\alpha_0, \beta_1, \cdots, \beta_{J-1})}^\top &= (D^\top D)^{-1} D^\top \mathbf{Y} &(7)\\
&= \frac{1}{I}\begin{pmatrix} J & \mathbf{J}_{J-1}^\top \\ \mathbf{J}_{J-1} & \mathbf{I}_{J-1}\end{pmatrix}^{-1} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I \mathbf{Y}_{-J,i}\end{pmatrix} &(8)\\
&= \frac{1}{I}\begin{pmatrix} 1 & -\mathbf{J}_{J-1}^\top \\ -\mathbf{J}_{J-1} & \mathbf{I}_{J-1} + \mathbf{J}_{J-1}\mathbf{J}_{J-1}^\top\end{pmatrix} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I \mathbf{Y}_{-J,i}\end{pmatrix} &(9)\\
&= \begin{pmatrix}\sum_{i=1}^I Y_{iJ}/I \\ (\sum_{i=1}^I \mathbf{Y}_{-J,i} - \sum_{i=1}^I Y_{iJ}\mathbf{J}_{J-1})/I\end{pmatrix} &(10)
\end{aligned}$$

iv. We replace β_J by $\sum_{i=1}^{J-1} -\beta_i$, then the design matrix $D = \begin{bmatrix}\mathbf{J}_{IJ} & \mathbf{J}_I \otimes \mathbf{L}_3\end{bmatrix}$, where $\mathbf{L}_3 = \begin{pmatrix}\mathbf{I}_{J-1}\\ \mathbf{J}_{J-1}^\top\end{pmatrix}$. So the estimation is:

$$\begin{aligned}
&\widehat{(\alpha_0, \beta_1, \cdots, \beta_{J-1})}^\top = (D^\top D)^{-1} D^\top \mathbf{Y} &(11)\\
=\ & \frac{1}{I}\begin{pmatrix} J & 0 \\ 0 & \mathbf{I}_{J-1} + \mathbf{J}_{J-1}\mathbf{J}_{J-1}^\top\end{pmatrix}^{-1} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I (\mathbf{Y}_{-J,i} - Y_{iJ}\mathbf{J}_{J-1})\end{pmatrix} &(12)\\
=\ & \frac{1}{I}\begin{pmatrix} J^{-1} & 0 \\ 0 & \mathbf{I}_{J-1} - \frac{1}{J}\mathbf{J}_{J-1}\mathbf{J}_{J-1}^\top\end{pmatrix} \begin{pmatrix}\sum_{i,j}^I Y_{ij} \\ \sum_{i=1}^I (\mathbf{Y}_{-J,i} - Y_{iJ}\mathbf{J}_{J-1})\end{pmatrix} &(13)\\
=\ & \begin{pmatrix}\sum_{i,j} Y_{ij}/(IJ) \\ \sum_{i=1}^I \mathbf{Y}_{-J,i}/I - \sum_{i,j} Y_{ij}\mathbf{J}_{J-1}/(IJ)\end{pmatrix} &(14)
\end{aligned}$$

3. *Proof.* In the slide of Lecture 5, we get that $\hat\beta_1 = (\mathbf{X}_1^\top \mathbf{X}_1)^{-1}\mathbf{X}_1^\top(\mathbf{y} - \mathbf{X}_2\hat\beta_2)$. And $\hat\beta_2$ is $(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{V}\mathbf{y}$, where $\mathbf{U} = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top)\mathbf{X}_2$, $\mathbf{V} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top$.

   If $\mathbf{X}_1, \mathbf{X}_2$ are orthogonal, then $\mathbf{X}_1^\top\mathbf{X}_2 = \mathbf{0}; \mathbf{X}_2^\top\mathbf{X}_1 = \mathbf{0}$. Therefore, $\mathbf{U} = \mathbf{X}_2$ and $\mathbf{U}^\top\mathbf{V} = \mathbf{X}_2$, so $\hat\beta_2 = (\mathbf{X}_2^\top\mathbf{X}_2)^{-1}\mathbf{X}_2\mathbf{y}$, and then $\hat\beta_1 = (\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top\mathbf{y}$. So $\hat\beta_1$ is not depend on $\mathbf{X}_2$ and $\hat\beta_2$ is not depend on $\mathbf{X}_1$. $\square$

4. A. The projection matrix in new model is:

$$H = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top = \mathbf{X}\mathbf{F}(\mathbf{F}^\top\mathbf{X}^\top\mathbf{X}\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{X}^\top = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

which is the same with it of design matrix $\mathbf{X}$. Therefore, the two models are equivalent in the sense of the projection procedures are the same. (which means the fitted values $\hat{\mathbf{y}}$ are the same, and along with result in B., if you get an estimation of the slope of one model, you can simutaneously get the other by a linear transform)

B. *Proof.*

$$\hat{\tilde{\beta}} = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\mathbf{y} = (\mathbf{F}^\top\mathbf{X}^\top\mathbf{X}\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{X}^\top\mathbf{y} = \mathbf{F}^{-1}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{F}^{-1}\hat{\beta}$$

□

C. Without loss of generality, we can suppose the first regressor is temperature. First, we have the transform formula:

$$T_{(^\circ F)} = T_{(^\circ C)} \times \frac{9}{5} + 32$$

Suppose $\mathbf{C}$ is the design matrix with temperature in Celsius ($\mathbf{T}_C$) and $\mathbf{F}$ is the design matrix with temperature in Fahrenheit ($\mathbf{T}_F$). Then denote $(\alpha^C, \beta_1^C, \beta_2^C, \cdots, \beta_p^C)^\top$ are the parameters of temperature in Celsius. $(\alpha^F, \beta_1^F, \beta_2^F, \cdots, \beta_p^F)^\top$ are the parameters of temperature in Fahrenheit. We have two models:

$$\mathbf{y} = \begin{bmatrix}\mathbf{J}_n & \mathbf{T}_C & \cdots\end{bmatrix}\begin{pmatrix}\alpha^C \\ \beta_1^C \\ \beta_2^C \\ \vdots \\ \beta_p^C\end{pmatrix} + \boldsymbol{\epsilon}_1$$

and

$$\mathbf{y} = \begin{bmatrix}\mathbf{J}_n & \mathbf{T}_F & \cdots\end{bmatrix}\begin{pmatrix}\alpha^F \\ \beta_1^F \\ \beta_2^F \\ \vdots \\ \beta_p^F\end{pmatrix} + \boldsymbol{\epsilon}_2 \tag{15}$$

$$= \begin{bmatrix}\mathbf{J}_n & \frac{9}{5}\mathbf{T}_C + 32 & \cdots\end{bmatrix}\begin{pmatrix}\alpha^F \\ \beta_1^F \\ \beta_2^F \\ \vdots \\ \beta_p^F\end{pmatrix} + \boldsymbol{\epsilon}_2 \tag{16}$$

$$= \begin{bmatrix}\mathbf{J}_n & \mathbf{T}_C & \cdots\end{bmatrix}\begin{pmatrix}\alpha^F + 32\beta_1^F \\ \frac{9}{5}\beta_1^F \\ \beta_2^F \\ \vdots \\ \beta_p^F\end{pmatrix} + \boldsymbol{\epsilon}_2 \tag{17}$$

So we have that $\hat{\beta}_1^C = \frac{9}{5}\hat{\beta}_1^F$ and $\hat{\beta}_i^C = \hat{\beta}_i^F$, for all $i > 1$.

5. A. *Proof.* First, since P_i are projection matrix, then they are symmetric and idempotent. $\forall$ vecotr $\mathbf{y}$, $\mathbf{y}^\top \mathbf{P}_i \mathbf{y} = \mathbf{y}^\top \mathbf{P}_i \mathbf{P}_i \mathbf{y} = \mathbf{y}^\top \mathbf{P}_i^\top \mathbf{P}_i \mathbf{y} = ||\mathbf{P}_i \mathbf{y}||^2 \geq 0$. So $\mathbf{P}_i$ are positive semidefinite. For every $\mathbf{y}$, we have $\mathbf{P}_1(\mathbf{I}-\mathbf{P}_1)\mathbf{y} = 0$, and $\mathbf{P}_1 - \mathbf{P}_2, \mathbf{P}_i$ are positive semidefinite. Therefore we have:

$$0 \leq [(\mathbf{I}-\mathbf{P}_1)\mathbf{y}]^\top (\mathbf{P}_1 - \mathbf{P}_2)[(\mathbf{I}-\mathbf{P}_1)\mathbf{y}] = -[(\mathbf{I}-\mathbf{P}_1)\mathbf{y}]^\top \mathbf{P}_2[(\mathbf{I}-\mathbf{P}_1)\mathbf{y}] \leq 0$$

So for every $\mathbf{y}$, we have $\mathbf{y}^\top(\mathbf{I}-\mathbf{P}_1)^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1)\mathbf{y} = 0$, then since $(\mathbf{I}-\mathbf{P}_1)^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1)$ is symmetric, we have $(\mathbf{I}-\mathbf{P}_1)^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1) = \mathbf{0}$.
Because:

$$(\mathbf{I}-\mathbf{P}_1)^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1) = (\mathbf{I}-\mathbf{P}_1)^\top \mathbf{P}_2^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1) = (\mathbf{P}_2(\mathbf{I}-\mathbf{P}_1))^\top \mathbf{P}_2(\mathbf{I}-\mathbf{P}_1) = \mathbf{0}$$

We have $\mathbf{P}_2(\mathbf{I}-\mathbf{P}_1) = \mathbf{0}$, which shows $\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1$. Meanwhile we have

$$\mathbf{P}_2 = \mathbf{P}_2^\top = (\mathbf{P}_2\mathbf{P}_1)^\top = \mathbf{P}_1^\top \mathbf{P}_2^\top = \mathbf{P}_1\mathbf{P}_1$$

Therefore $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_2$. □

B. *Proof.* Use results in A. We have:

$$(\mathbf{P}_1 - \mathbf{P}_2)^2 = \mathbf{P}_1^2 - \mathbf{P}_1\mathbf{P}_2 - \mathbf{P}_2\mathbf{P}_1 + \mathbf{P}_2^2 = \mathbf{P}_1 + \mathbf{P}_2 - \mathbf{P}_2 - \mathbf{P}_2 = \mathbf{P}_1 - \mathbf{P}_2$$

Therefore $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix. □