BST 140.751
Problem Set 1
Due: September 12, 2017.

1 Vector spaces and inner products

1. Let $\mathbf{W}$ be a subspace of $\mathbb{R}^n$, and $\mathbf{W}^\perp$ its orthogonal complement. Show that if the dimension of $\mathbf{W}$ is k, then the dimension of $\mathbf{W}^\perp$ is $n-k$.

2. Let $\mathbf{V}$ be an inner product space and $\mathbf{u}, \mathbf{v} \in \mathbf{V}$.

   (a) Show that $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}||\dot{|}|\mathbf{v}||$

   (b) Show that $2||\mathbf{u}||^2 + 2||\mathbf{v}||^2 = ||\mathbf{u} + |\mathbf{v}||^2 + ||\mathbf{u} - \mathbf{v}||^2$.

3. Let us denote the projection of $\mathbf{y}$ on $\mathbf{x}$ as $\Pi(\mathbf{y}|\mathbf{x})$. Show that $\mathbf{y} - \Pi(\mathbf{y}|\mathbf{x})$ is orthogonal to $\mathbf{x}$, for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

2 Regression

1. Let $\mathbf{y}$ and $\mathbf{x}$ be one dimensional vectors of length n. Give the relationship between the slope from regressing $\mathbf{y}$ on $\mathbf{x}$ and $\mathbf{x}$ on $\mathbf{y}$.

2. Consider the residuals after mean only regression. Argue that they sum to 0.

3. Consider the residuals after regression through the origin. Argue that they are orthogonal to the regressor.

4. Consider the residuals after regression through the origin. Argue that they need not sum to 0.

5. Consider the residuals from ordinary linear regression. Argue that the residuals are orthogonal to both $\mathbf{J}_n$ and $\mathbf{x}$.

3 Least squares

1. Show that $\mathbf{I} - \mathbf{H}$ is an idempotent matrix when $\mathbf{H}$ is idempotent.

2. Let $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2]$ be an $n \times 2$ design matrix and consider

$$||\mathbf{y} - \mathbf{X}\beta||^2$$

   where $\beta = (\beta_1 \ \beta_2)'$. Show that $\hat{\beta}_2$ can be obtained by taking the residuals after regressing $\mathbf{x}_1$ out of $\mathbf{y}$ and $\mathbf{x}_2$ then performing regression through the origin on the residuals.

3. Argue that $\mathbf{X}$, $\mathbf{X}'$, $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$ all have the same matrix rank.

4. Suppose that $\mathbf{X}$ is such that $\mathbf{X}'\mathbf{X} = \mathbf{I}$. Find the associated least squares estimate of β.

4 Computing and analysis

1. Write an R function called `mylm()` that takes the response vector $\mathbf{y}$ and the matrix of covariates $\mathbf{X}$ as input, and returns the following variables:

   - beta, the vector of least squares estimates,
   - fitted, the vector of fitted values,
   - residuals, the vector of residuals.

   To fit an intercept, the elements in the first column of $\mathbf{X}$ have to be equal to one, so your function should have an option to add a vector of ones to the matrix with the predictors.

2. Find a dataset to try out your function (you can simulate one if you like), and compare the results to the one from the `lm()` function.