Homework 5

Due Thursday, 9/24/20

This homework goes through the most important properties of the "hat" matrix. Throughout this problem set, the matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is assumed to be a non-random, full rank matrix with $p \leq n$. In simple linear regression, $p = 2$. We let $\boldsymbol{Y} \in \mathbb{R}^n$ be any vector and define

$$\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T, \quad \boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{H}.$$

I have also provided some definitions and properties below that will be useful throughout the problem set.

- Let $\boldsymbol{x} \in \mathbb{R}^n$. Then the **2-norm** $\|\cdot\|_2$ is defined as $(\|\boldsymbol{x}\|_2)^2 = \|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T\boldsymbol{x}$.
- Let $\boldsymbol{A} \in \mathbb{R}^{n\times d}$. The **image** of $\boldsymbol{A}$ is defined as

  $$\operatorname{Im}(\boldsymbol{A}) = \left\{\boldsymbol{A}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^d\right\} \subseteq \mathbb{R}^n$$

  This is also called the **column space** of $\boldsymbol{A}$ or the **span** of the columns of $\boldsymbol{A}$.
- Suppose $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ is defined as

  $$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

- The **trace** of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is defined as $\operatorname{Tr}(\boldsymbol{A}) = \sum_{i=1}^n \boldsymbol{A}_{ii}$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$, then $\operatorname{Tr}(\boldsymbol{A}) = \sum_{i=1}^n \lambda_i$. You may find this useful in problem 3(c).

1. Deriving $\boldsymbol{H}$ in ordinary least squares.

   (a) Let $\boldsymbol{B} \in \mathbb{R}^p$ and define

   $$f(\boldsymbol{B}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}).$$

   Use the properties on slide 12 of Lecture 8 to show that

   notes + stanford

   $$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{B}\in\mathbb{R}^p}{\arg\min} f(\boldsymbol{B}) = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

   (b) Let $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ be the predicted values and estimated residuals. Show that $\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}$ and $\hat{\boldsymbol{\epsilon}} = \boldsymbol{Q}\boldsymbol{Y}$.

   notes + stanford

2. The action of $\boldsymbol{H}$.

prove im(H) con im(x): let a random vector in im(H) => use brackets => it is in im(X)
prove im(X) con im(H): let a random vector in im(X) => it has form of X times a p*1 vector, this vector can be constructed by multiplying the p*n part of H with a n*1 random vector

   (a) Use the definition of $\boldsymbol{H}$ to show that $\text{Im}(\boldsymbol{H}) = \text{Im}(\boldsymbol{X})$.

   (b) Show that $\boldsymbol{H}$ and $\boldsymbol{Q}$ are symmetric, idempotent and orthogonal to one another. That is, show

      i. $\boldsymbol{H} = \boldsymbol{H}^T$ and $\boldsymbol{Q} = \boldsymbol{Q}^T$.

      HW7 - dsci

      ii. $\boldsymbol{H}^2 = \boldsymbol{H}$ and $\boldsymbol{Q}^2 = \boldsymbol{Q}$.

      iii. $\boldsymbol{HQ} = \boldsymbol{0}$.

   (c) Use parts (a) and (b) to show that $\boldsymbol{H}$ projects vectors in $\mathbb{R}^n$ onto the image of $\boldsymbol{X}$, AND is an **orthogonal projection matrix**. That is, prove the following:

      i. If $\boldsymbol{v} \in \mathbb{R}^n$, then $\boldsymbol{Hv} \in \text{Im}(\boldsymbol{X})$. If $\boldsymbol{u} \in \text{Im}(\boldsymbol{X})$, then $\boldsymbol{Hu} = \boldsymbol{u}$.

      ii. Let $\boldsymbol{v} \in \mathbb{R}^n$. Then $\boldsymbol{Hv}$ is the closest vector in $\text{Im}(\boldsymbol{X})$ to $\boldsymbol{v}$. That is,

$$\boldsymbol{Hv} = \underset{\boldsymbol{u} \in \text{Im}(\boldsymbol{X})}{\arg\min} \|\boldsymbol{v} - \boldsymbol{u}\|_2^2.$$

      (Hint: write $\boldsymbol{v} - \boldsymbol{u}$ as $\boldsymbol{v} - \boldsymbol{u} = \boldsymbol{H}(\boldsymbol{v} - \boldsymbol{u}) + \boldsymbol{Q}(\boldsymbol{v} - \boldsymbol{u})$ and expand $\|\boldsymbol{v} - \boldsymbol{u}\|_2^2$.)

3. The eigen-decomposition of $\boldsymbol{H}$ and $\boldsymbol{Q}$. For this part, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\boldsymbol{H}$. Note a consequence of 2(a) is $\text{rank}(\boldsymbol{H}) = \text{rank}(\boldsymbol{X})$.

   (a) Use the fact that $\boldsymbol{H}$ is idempotent to show that $\lambda_i = 0$ or $\lambda_i = 1$ for all $i = 1, \ldots, n$. In terms of n and p, how many eigenvalues are 1 and how many are 0?

   (b) If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{H}$, what are the eigenvalues of $\boldsymbol{Q}$?

   (c) Use parts (a) and (b) to show that $\text{Tr}(\boldsymbol{H}) = p$ and $\text{Tr}(\boldsymbol{Q}) = n - p$.

4. Now suppose $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ for some constant $\boldsymbol{\beta} \in \mathbb{R}^p$, where $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$. Note that in simple linear regression, $p = 2$.

   (a) Use the definition of $\hat{\boldsymbol{\beta}}$, derived in 1(a), to show that $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

   (b) Recall that for two random vectors $\boldsymbol{W}, \boldsymbol{Z} \in \mathbb{R}^n$,

$$\text{Cov}(\boldsymbol{W}, \boldsymbol{Z}) = \mathbb{E}\left[\{\boldsymbol{W} - \mathbb{E}(\boldsymbol{W})\}\{\boldsymbol{Z} - \mathbb{E}(\boldsymbol{Z})\}^T\right] \in \mathbb{R}^{n \times n}.$$

   Use 2(b) to show that $\text{Cov}\left(\hat{\boldsymbol{Y}}, \hat{\boldsymbol{\epsilon}}\right) = \boldsymbol{0}$.

   (c) **PhD problem**: Does 3(b) hold if $\text{Var}(\boldsymbol{\epsilon})$ is not a multiple of the identity matrix?

   (d) **PhD problem**: Now assume $\boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n\right)$.

      (i) Show that part (b) implies $(\hat{\boldsymbol{Y}}, \hat{\boldsymbol{\beta}})$ is independent of $\hat{\boldsymbol{\epsilon}}$

      (ii) Is (i) true if $\text{Var}(\boldsymbol{\epsilon})$ is not a multiple of the identity matrix?

      (iii) If $\boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n\right)$, what would you expect to see if you plotted $\hat{\boldsymbol{Y}}$ vs. $\hat{\boldsymbol{\epsilon}}$? Explain.