STAT151A Homework 2.

Your name here

This homework is due on Gradescope on **Monday September 30th at 9pm**.

1 Ordinary least squares in matrix form

Consider simple least squares regression $y_n = \beta_1 + \beta_2 x_n + \varepsilon_n$, where x_n is a scalar. Assume that we have N datapoints. We showed directly that the least–squares solution is given by

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x} \quad \textit{and} \qquad \hat{\beta}_2 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{xx} - \overline{x}^2}.$$

Let us re–derive this using matrix notation.

(a)

Write simple linear regression in the form $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Be precise about what goes into each entry of $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$. What are the dimensions of each?

(b)

We proved that the optimal $\hat{\boldsymbol{\beta}}$ satisfies $\boldsymbol{X}^\intercal \boldsymbol{X} \hat{\boldsymbol{\beta}} = \boldsymbol{X}^\intercal \boldsymbol{Y}$. Define the "barred quantities"

$$\begin{aligned}
\overline{y} &= \frac{1}{N}\sum_{n=1}^{N} y_n \\
\overline{x} &= \frac{1}{N}\sum_{n=1}^{N} x_n \\
\overline{xy} &= \frac{1}{N}\sum_{n=1}^{N} x_n y_n \\
\overline{xx} &= \frac{1}{N}\sum_{n=1}^{N} x_n^2,
\end{aligned}$$

In terms of the barred quantities and the number of datpoints N, write expressions for $\boldsymbol{X}^\intercal \boldsymbol{X}$ and $\boldsymbol{X}^\intercal \boldsymbol{Y}$.

(c)

When is $\boldsymbol{X}^\intercal \boldsymbol{X}$ invertible? Write a formal expression in terms of the barred quantities. Interpret this condition intuitively in terms of the distribution of the regressors x_n.

(d)

Using the formula for the inverse of a 2×2 matrix, find an expression for $\hat{\boldsymbol{\beta}}$, and confirm that we get the same answer that we got by solving directly.

(e)

In the case where $\boldsymbol{X}^\intercal \boldsymbol{X}$ is *not* invertible, find three distinct values of $\boldsymbol{\beta}$ that all achieve the same sum of squared residuals $\boldsymbol{\varepsilon}^\intercal \boldsymbol{\varepsilon}$.

2 Probability and matrices

For this problem, assume that $y_n = \beta_0 + x_n \beta_1 + \varepsilon_n$ for scalar x_n and some fixed β_0 and β_1. Assume that

- The residuals ε_n are IID with $\mathbb{E}\left[\varepsilon_n\right] = 0$ and $\mathrm{Var}\left(\varepsilon_n\right) = \sigma^2$.
- The regressors x_n are IID with $\mathbb{E}\left[x_n\right] = \mu$ and $\mathrm{Var}\left(x_n\right) = \nu^2$.
- The residuals are independent of the regressors.

(a)

Evaluate the following expressions. (You may need to remind yourself of the definition of conditional expectation and variance.)

- $\mathbb{E}\left[y_n\right]$
- $\mathrm{Var}\left(y_n\right)$
- $\mathbb{E}\left[y_n | x_n\right]$
- $\mathrm{Var}\left(y_n | x_n\right)$

(b)

Compute the following limits using the LLN, or say that the limit does not exist or is infinite.

- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \varepsilon_n$
- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N \varepsilon_n^2$
- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N x_n$
- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N x_n^2$
- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N x_n \varepsilon_n$
- $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N x_n y_n$

(c)

Compute the following limits using the CLT, or say that the limit does not exist or is infinite.

- $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^N \varepsilon_n$
- $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^N y_n$
- $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^N (y_n - (\beta_0 + x_n \beta_1))$

(d)

Noting that this is simple linear regression, let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{\varepsilon}$ be as in the solution to question one above. Evaluate the following limits, or say that the limit does not exist or is infinite.

Here, $(\boldsymbol{A})_{ij}$ denotes the i,j–th entry of the matrix $\boldsymbol{A}$. Let the regressor x_n be in the second column of $\boldsymbol{X}$.

- $\lim_{N\to\infty} \frac{1}{N} \mathbf{1}^\intercal \boldsymbol{\varepsilon}$
- $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \mathbf{1}^\intercal \boldsymbol{\varepsilon}$
- $\lim_{N\to\infty} \frac{1}{N} (\boldsymbol{X}^\intercal \boldsymbol{X})_{11}$
- $\lim_{N\to\infty} \frac{1}{N} (\boldsymbol{X}^\intercal \boldsymbol{X})_{12}$
- $\lim_{N\to\infty} \frac{1}{N} (\boldsymbol{X}^\intercal \boldsymbol{X})_{22}$
- $\lim_{N\to\infty} (\boldsymbol{X}^\intercal \boldsymbol{X})_{11}$
- $\lim_{N\to\infty} \frac{1}{N} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$
- $\lim_{N\to\infty} \frac{1}{\sqrt{N}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\intercal (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$

Hint: Write the matrix expressions as sums over $n = 1$ to N.

3 One-hot encoding

Consider a one–hot encoding of a variable z_n that takes three distinct values, "a", "b", and "c". That is, let

$$\boldsymbol{x}_n = \begin{cases} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & \text{when } z_n = a \\ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} & \text{when } z_n = b \\ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} & \text{when } z_n = c \end{cases}$$

Let $\boldsymbol{X}$ be the regressor matrix with $\boldsymbol{x}_n^\intercal$ in row n.

(a)

Let N_a be the number of observations with $z_n =$ a, and let $\sum_{n:z_n=a}$ denote a sum over rows with $z_n =$ a, with analogous definitions for b and c. In terms of these quantities, write expressions for $\boldsymbol{X}^\intercal\boldsymbol{X}$ and $\boldsymbol{X}^\intercal\boldsymbol{Y}$.

(b)

When is $\boldsymbol{X}^\intercal\boldsymbol{X}$ invertible? Explain intuitively why the regression problem cannot be solved when $\boldsymbol{X}^\intercal\boldsymbol{X}$ is not invertible. Write an explicit expression for $(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}$ when it is invertible.

(c)

Using your previous answer, show that the least squares vector $\hat{\boldsymbol{\beta}}$ is the mean of y_n within distinct values of z_n.

(e)

Suppose now you include a constant in the regression, so that

$$y_n = \alpha + \boldsymbol{\beta}^\intercal \boldsymbol{x}_n + \varepsilon_n,$$

and let $\boldsymbol{X}'$ denote the regressor matrix for this regression with coefficient vector $(\alpha, \boldsymbol{\beta}^\intercal)^\intercal$. Write an expression for $\boldsymbol{X}'^\intercal \boldsymbol{X}'$ and show that it is not invertible.

(f)

Find three distinct values of $(\alpha, \boldsymbol{\beta}^\intercal)$ that all give the exact same fit $\alpha + \boldsymbol{\beta}^\intercal \boldsymbol{x}_n$.

4 Correlated regressors

Suppose that $y_n = \boldsymbol{x}_n^\intercal \boldsymbol{\beta} + \varepsilon_n$ for some $\boldsymbol{\beta}$. Suppose that $\mathbb{E}\left[\varepsilon_n\right] = 0$ and $\mathrm{Var}\left(\varepsilon_n\right) = \sigma^2$, and ε_n are independent of each other and the $\boldsymbol{x}_n$.

Let $\boldsymbol{x}_n \in \mathbb{R}^2$, where

- $\boldsymbol{x}_n$ is independent of $\boldsymbol{x}_m$ for $n \neq m$,
- $\mathbb{E}\left[\boldsymbol{x}_{n1}\right] = \mathbb{E}\left[\boldsymbol{x}_{n2}\right] = 0$,
- $\mathrm{Var}\left(\boldsymbol{x}_{n1}\right) = \mathrm{Var}\left(\boldsymbol{x}_{n2}\right) = 1$, and
- $\mathbb{E}\left[\boldsymbol{x}_{n1} \boldsymbol{x}_{n2}\right] = \rho$.

(a)

If $|\rho| < 1$, is $\boldsymbol{X}^\intercal \boldsymbol{X}$ always, sometimes, or never invertible?

(b)

If $|\rho| = 1$, is $\boldsymbol{X}^\intercal \boldsymbol{X}$ always, sometimes, or never invertible?

(c)

What is $\lim_{N \to \infty} \frac{1}{N} \boldsymbol{X}^\intercal \boldsymbol{X}$? When is the limit invertible?

(d)

State intuitively why there is no unique $\hat{\boldsymbol{\beta}}$ when $\rho = 1$. When $\rho = 1$, give two distinct values of $\boldsymbol{\beta}$ that result in the same fit $\boldsymbol{\beta}^\intercal \boldsymbol{x}_n$.

5 Matrix square roots

In the last homework, we proved that if $\boldsymbol{A}$ is a square symmetric matrix with eigenvalues $\boldsymbol{u}_p$ and eigenvectors λ_p, then we can write $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^\intercal$, where $\boldsymbol{U} = (\boldsymbol{u}_1 \ldots \boldsymbol{u}_p)$ has $\boldsymbol{u}_p$ in its p–th column, and Λ is diagonal with λ_p in the p–th diagonal entry. We also have that the eigenvectors can be taken to be orthonormal without loss of generality.

We will additionally assume that $\boldsymbol{A} = \boldsymbol{X}^\intercal\boldsymbol{X}$ for some (possibly non–square) matrix $\boldsymbol{X}$.

Define $\Lambda^{1/2}$ to be the diagonal matrix with $\sqrt{\lambda_p}$ on the p–th diagonal.

- Prove that, since $\boldsymbol{A} = \boldsymbol{X}^\intercal\boldsymbol{X}$, $\lambda_p \geq 0$, and so $\Lambda^{1/2}$ is always real–valued.
  When the eignevalues are non–negative, we say that $\boldsymbol{A}$ is "positive semi–definite." (Hint: using the fact that $\lambda_p = \boldsymbol{u}_p^\intercal\boldsymbol{A}\boldsymbol{u}_p$, show that λ_p is the square of something.)
- Show that if we take $\boldsymbol{Q} = \boldsymbol{U}\Lambda^{1/2}\boldsymbol{U}^\intercal$ then $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}^\intercal$. We say that $\boldsymbol{Q}$ is a "matrix square root" of $\boldsymbol{A}$.
- Show that we also have $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}$ (without the second transpose).
- Show that, if $\boldsymbol{V}$ is any orthonormal matrix (a matrix with orthonormal columns), then $\boldsymbol{Q}' = \boldsymbol{Q}\boldsymbol{V} \neq \boldsymbol{Q}$ also satisfies $\boldsymbol{A} = \boldsymbol{Q}'\boldsymbol{Q}'^\intercal$. This shows that the matrix square root is not unique. (This fact can be thought of as the matrix analogue of the fact that $4 = 2 \cdot 2$ but also $4 = (-2) \cdot (-2)$).
- Show that if $\lambda_p > 0$ then $\boldsymbol{Q}$ is invertible.
- Show that, if $\boldsymbol{Q}$ is invertible, then the columns of $\boldsymbol{X}\boldsymbol{Q}^{-1}$ are orthonormal. (Hint: show that $(\boldsymbol{X}\boldsymbol{Q}^{-1})^\intercal(\boldsymbol{X}\boldsymbol{Q}^{-1})$ is the identity matrix.)