

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

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

Let us re-derive this using matrix notation.

(a)

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

(b)

We proved that the optimal $\hat{\boldsymbol{\beta}}$ satisfies $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{Y}$. Define the “barred quantities”

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

In terms of the barred quantities and the number of datapoints N , write expressions for $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{X}^\top \mathbf{Y}$.

(c)

When is $\mathbf{X}^\top \mathbf{X}$ 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 $\hat{\beta}$, and confirm that we get the same answer that we got by solving directly.

(e)

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

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 $\mathbb{E}[\varepsilon_n] = 0$ and $\text{Var}(\varepsilon_n) = \sigma^2$.
- The regressors x_n are IID with $\mathbb{E}[x_n] = \mu$ and $\text{Var}(x_n) = \nu^2$.
- 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.)

- $\mathbb{E}[y_n]$
- $\text{Var}(y_n)$
- $\mathbb{E}[y_n | x_n]$
- $\text{Var}(y_n | x_n)$

(b)

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

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

(c)

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

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

(d)

Noting that this is simple linear regression, let \mathbf{X} , \mathbf{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, $(\mathbf{A})_{ij}$ denotes the i, j -th entry of the matrix \mathbf{A} . Let the regressor x_n be in the second column of \mathbf{X} .

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

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

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

Let \mathbf{X} be the regressor matrix with \mathbf{x}_n^\top 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 $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{X}^\top \mathbf{Y}$.

(b)

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

(c)

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

(e)

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

$$y_n = \alpha + \beta^\top \mathbf{x}_n + \varepsilon_n,$$

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

(f)

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

4 Correlated regressors

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

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

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

(a)

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

(b)

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

(c)

What is $\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}^\top \mathbf{X}$? When is the limit invertible?

(d)

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

5 Matrix square roots

In the last homework, we proved that if \mathbf{A} is a square symmetric matrix with eigenvalues λ_p and eigenvectors \mathbf{u}_p , then we can write $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U} = (\mathbf{u}_1 \dots \mathbf{u}_p)$ has \mathbf{u}_p in its p -th column, and $\mathbf{\Lambda}$ 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 $\mathbf{A} = \mathbf{X}^\top \mathbf{X}$ for some (possibly non-square) matrix \mathbf{X} .

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

- Prove that, since $\mathbf{A} = \mathbf{X}^\top \mathbf{X}$, $\lambda_p \geq 0$, and so $\mathbf{\Lambda}^{1/2}$ is always real-valued. When the eigenvalues are non-negative, we say that \mathbf{A} is “positive semi-definite.” (Hint: using the fact that $\lambda_p = \mathbf{u}_p^\top \mathbf{A} \mathbf{u}_p$, show that λ_p is the square of something.)
- Show that if we take $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top$ then $\mathbf{A} = \mathbf{Q}\mathbf{Q}^\top$. We say that \mathbf{Q} is a “matrix square root” of \mathbf{A} .
- Show that we also have $\mathbf{A} = \mathbf{Q}\mathbf{Q}$ (without the second transpose).
- Show that, if \mathbf{V} is any orthonormal matrix (a matrix with orthonormal columns), then $\mathbf{Q}' = \mathbf{Q}\mathbf{V} \neq \mathbf{Q}$ also satisfies $\mathbf{A} = \mathbf{Q}'\mathbf{Q}'^\top$. 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 \mathbf{Q} is invertible.
- Show that, if \mathbf{Q} is invertible, then the columns of $\mathbf{X}\mathbf{Q}^{-1}$ are orthonormal. (Hint: show that $(\mathbf{X}\mathbf{Q}^{-1})^\top(\mathbf{X}\mathbf{Q}^{-1})$ is the identity matrix.)