

STAT151A Homework 6: Due April 19th

Your name here

1 Fit and regressors

Given a regression on \mathbf{X} with P regressors, and the corresponding \mathbf{Y} , $\hat{\mathbf{Y}}$, and $\hat{\varepsilon}$, define the following quantities:

$$RSS := \hat{\varepsilon}^\top \hat{\varepsilon} \quad (\text{Residual sum of squares})$$

$$TSS := \mathbf{Y}^\top \mathbf{Y} \quad (\text{Total sum of squares})$$

$$ESS := \hat{\mathbf{Y}}^\top \hat{\mathbf{Y}} \quad (\text{Explained sum of squares})$$

$$R^2 := \frac{ESS}{TSS}.$$

- Prove that $RSS + ESS = TSS$.
- Express R^2 in terms of TSS and RSS .
- What is R^2 when we include no regressors? ($P = 0$)
- What is R^2 when we include N linearly independent regressors? ($P = N$)
- Can R^2 ever decrease when we add a regressor? If so, how?
- Can R^2 ever stay the same when we add a regressor? If so, how?
- Can R^2 ever increase when we add a regressor? If so, how?
- Does a high R^2 mean the regression is correctly specified? Why or why not?
- Does a low R^2 mean the regression is incorrectly specified? Why or why not?

The next questions will be about the F-test statistic for the null $H_0 : \beta = \mathbf{0}$,

$$\phi = \hat{\beta}^\top (\mathbf{X}^\top \mathbf{X}) \hat{\beta} / (P \hat{\sigma}^2)$$

- Write the F-test statistic ϕ in terms of TSS and RSS , and P .
- Can ϕ ever decrease when we add a regressor? If so, how?
- Can ϕ ever stay the same when we add a regressor? If so, how?
- Can ϕ ever increase when we add a regressor? If so, how?

2 Omitted variable bias

For this problem, let $(\mathbf{x}_n, \mathbf{z}_n, y_n)$ be IID random variables, where $\mathbf{x}_n \in \mathbb{R}^{P_X}$ and $\mathbf{z}_n \in \mathbb{R}^{P_Z}$. Suppose that \mathbf{x}_n and \mathbf{z}_n are uncorrelated, so that $\mathbb{E}[\mathbf{x}_n \mathbf{z}_n^\top] = \mathbf{0}$.

Let $y_n = \mathbf{x}_n^\top \beta + \mathbf{z}_n^\top \gamma + \varepsilon_n$, where ε_n is mean zero, unit variance, and independent of \mathbf{x}_n and \mathbf{z}_n .

a

Take $P_X = P_Z = 1$ (i.e. scalar regressors). Show that there exists x_n and z_n such that $\mathbb{E}[x_n z_n] = 0$ but $\mathbb{E}[z_n | x_n] \neq 0$ for some x_n . (A single counterexample will be enough.)

b

Now return to the general case. Let $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ denote the OLS estimator from the regression on \mathbf{X} alone. Derive an expression for $\mathbb{E}[\hat{\beta}]$, where the expectation is taken over \mathbf{X} , \mathbf{Y} , and \mathbf{Z} .

Hint: by the Tower property,

$$\mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Z}] = \mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\mathbf{Z} | \mathbf{X}]].$$

c

Using (b), derive an expression for the bias for a fixed \mathbf{x}_{new} , i.e.

$$\mathbb{E}[y_{\text{new}} - \mathbf{x}_{\text{new}}^\top \hat{\beta} | \mathbf{x}_{\text{new}}],$$

in terms of β , γ , and the conditional expectation $\mathbb{E}[\mathbf{z}_{\text{new}} | \mathbf{x}_{\text{new}}]$.

d

Using your result from (c), show that the predictions are biased at \mathbf{x}_{new} when omitting the variables \mathbf{z}_n from the regression precisely when $\gamma^\top \mathbb{E}[\mathbf{z}_n | \mathbf{x}_n] \neq 0$. Using your result from (a), show that this bias can be expected to occur in general — that is, omitting variables can often induce biased predictions at a point.

3 Estimating leave-one-out CV

This homework problem derives a closed-form estimate of the leave-one-out cross-validation error for regression. We will use the Sherman-Woodbury formula. Let A denote an invertible matrix, and \mathbf{u} and \mathbf{v} vectors the same length as A . Then

$$(A + \mathbf{u}\mathbf{v}^\top)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}A^{-1}}{1 + \mathbf{v}^\top A^{-1}\mathbf{u}}.$$

(For reference, I provide a proof in the notes for lecture 21.)

We will also use the following definition of a “leverage score,” $h_n := \mathbf{x}_n^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_n$. We will discuss leverage scores more in the last lecture, but for now it’s enough that you know what it is. Note that $h_n = (\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)_{nn}$ is the n -th diagonal entry of the projection matrix $\frac{\mathbf{P}}{\mathbf{X}}$.

Let $\hat{\boldsymbol{\beta}}_{-n}$ denote the estimate of $\hat{\boldsymbol{\beta}}$ with the datapoint n left out. For leave-one-out CV, we want to estimate

$$MSE_{LOO} := \frac{1}{N} \sum_{n=1}^N (y_n - \mathbf{x}_n^\top \hat{\boldsymbol{\beta}}_{-n})^2.$$

Note that doing so naively requires computing N different regressions. We will derive a much more efficient formula.

Let \mathbf{X}_{-n} denote the \mathbf{X} matrix with row n left out, and \mathbf{Y}_{-n} denote the \mathbf{Y} matrix with row n left out.

a

Prove that

$$\hat{\boldsymbol{\beta}}_{-n} = (\mathbf{X}_{-n}^\top \mathbf{X}_{-n})^{-1} \mathbf{X}_{-n}^\top \mathbf{Y}_{-n} = (\mathbf{X}^\top \mathbf{X} - \mathbf{x}_n \mathbf{x}_n^\top)^{-1} (\mathbf{X}^\top \mathbf{Y} - \mathbf{x}_n y_n)$$

b

Using the Sherman-Woodbury formula, derive the following expression:

$$(\mathbf{X}^\top \mathbf{X} - \mathbf{x}_n \mathbf{x}_n^\top)^{-1} = (\mathbf{X}^\top \mathbf{X})^{-1} + \frac{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_n \mathbf{x}_n^\top (\mathbf{X}^\top \mathbf{X})^{-1}}{1 - h_n}$$

c

Combine (a) and (b) to derive the following explicit expression for $\hat{\beta}_{-n}$:

$$\hat{\beta}_{-n} = \hat{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_n \frac{h_n}{1 - h_n} \hat{\varepsilon}_n$$

d

Using (c), derive the following explicit expression the leave-one-out error on the n -th observation:

$$y_n - \mathbf{x}_n^\top \hat{\beta}_{-n} = \frac{\hat{\varepsilon}_n}{1 - h_n}.$$

e

Using (d), prove that

$$MSE_{LOO} := \frac{1}{N} \sum_{n=1}^N \frac{\hat{\varepsilon}_n^2}{(1 - h_n)^2},$$

where $\hat{\varepsilon}_n = y_n - \hat{y}_n$ is the residual from the full regression without leaving any data out. Using this formula, MSE_{LOO} can be computed using only the original regression and $(\mathbf{X}^\top \mathbf{X})^{-1}$.

f

Prove that $\sum_{n=1}^N h_n = N - P$, and $0 \leq h_n \leq 1$. Hint: if \mathbf{v} is a vector with a 1 in entry n and 0 otherwise, then $h_n = \mathbf{v}^\top \mathbf{P}_X \mathbf{v}$, and projection cannot increase a vector's norm. Recall also that $\text{trace}\left(\mathbf{P}_X\right) = N - P$.

g

Using (e) and (f), prove that $MSE_{LOO} > RSS = \frac{1}{N} \sum_{n=1}^N \hat{\varepsilon}_n^2$. That is, the RSS underestimates the leave-one-out cross-validation error.