Econ 2120: Section 1

Part I - Orthogonality and Projection

Ashesh Rambachan

Fall 2018

Outline

Outline

Inner Products

For a vector space $\mathcal{V}$, an **inner product** $\langle \cdot, \cdot \rangle$ is a function defined on $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ that satisfies the following properties:

<u>Symmetry</u>: For all $v_1, v_2 \in \mathcal{V}$,

$$\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle.$$

<u>Linearity</u>: For all $\alpha \in \mathbb{R}$ and $v_1, v_2, v_3 \in \mathcal{V}$

$$\langle \alpha v_1 + v_2, v_3 \rangle = \alpha \langle v_1, v_3 \rangle + \langle v_2, v_3 \rangle.$$

<u>Positive-definiteness</u>: For all $v_1 \in \mathcal{V}$,

$$\langle v_1, v_1 \rangle \geq 0$$

with equality if and only if $v_1 = 0$.

Inner Products

v_1, v_2 are **orthogonal** if

$$\langle v_1, v_2 \rangle = 0.$$

Write $v_1 \perp v_2$.

Consider a subspace $X \subset \mathcal{V}$. v_1 is **orthogonal** to X if

$$\langle v_1, x \rangle = 0$$

for all $x \in \mathcal{X}$. Write $v_1 \perp X$.

Norms

Next, the **norm** associated with $\langle\cdot,\cdot\rangle$ is

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

The norm satisfies the following properties for all $v \in \mathcal{V}$:

Positivity: $\|v\| \geq 0$ with $\|v\| = 0$ if and only if $v = 0$.

Homogeneity: For all $\alpha \in \mathbb{R}$, $\|\alpha \cdot v\| = |\alpha|\|v\|$.

Triangle Inequality: For all $v_1, v_2 \in \mathcal{V}$,

$$\|x + y\| \leq \|x\| + \|y\|.$$

Cauchy-Schwarz Inequality

Theorem (Cauchy-Schwarz)

For all $v_1, v_2 \in \mathcal{V}$,

$$|\langle v_1, v_2 \rangle| \leq \|v_1\| \|v_2\|.$$

Projection Theorem

Theorem (Projection)

Let $\mathcal{V}$ be a vector space, let $X \subset \mathcal{V}$ be a subspace and fix $y \in \mathcal{V}$. Then,

$$x^* = \arg\min_{x \in \mathcal{X}} \|y - x\|$$

if and only if

$$\langle y - x^*, x \rangle = 0$$

for all $x \in \mathcal{X}$.

Proof: ($\Leftarrow$)

Consider $x^* \in \mathcal{X}$ such that for all $x \in \mathcal{X}$,

$$\langle y - x^*, x \rangle = 0.$$

$\forall x \in \mathcal{X}$, define $x' \in \mathcal{X}$ and $t \in \mathbb{R}_+$ such that $\|x'\| = 1$ and

$$x = x^* + t \cdot x'.$$

Let $t = \|x - x^*\|$ and $x' = (x - x^*)/t$

Then,

$$\begin{aligned}
\|y - x\|^2 &= \|y - x^* - t \cdot x'\|^2 \\
&= \langle (y - x^*) - t \cdot x', (y - x^*) - t \cdot x' \rangle \\
&= \|y - x^*\|^2 - 2t\langle y - x^*, x' \rangle + t^2 \geq \|y - x^*\|^2
\end{aligned}$$

with equality if and only if $t = 0$ or $x = x^*$.

Proof: ($\Rightarrow$)

Consider $x^* \in \mathcal{X}$ such that

$$x^* = \arg\min_{x \in \mathcal{X}} \|y - x\|.$$

Suppose FSOC there exists an $x \in \mathcal{X}$ such that

$$\langle y - x^*, x, \neq \rangle 0.$$

WLOG, suppose that $\|x\| = 1$ and call $\langle y - x^*, x \rangle = \delta$. Let $x_1 = x^* + \delta \cdot x$. Then,

$$\begin{aligned} \|y - x_1\|^2 &= \langle y - x^* - \delta \cdot x, y - x^* - \delta \cdot x \rangle \\ &= \|y - x^*\|^2 - \langle y - x^*, \delta \cdot x \rangle - \langle \delta \cdot x, y - x^* \rangle + \delta^2 \\ &= \|y - x^*\|^2 - \delta^2 < \|y - x^*\|^2. \end{aligned}$$

Note: For this direction, need to assume $\mathcal{V}$ is a Hilbert space and X is a closed subspace of $\mathcal{X}$ for x^* to be guaranteed to exist. Will be true for $\mathcal{V} = \mathbb{R}^d$ with the Euclidean inner product.

Exercise 1

Let H be a vector space with an inner product $\langle\cdot,\cdot\rangle$.

(1) Given $f, g \in H$, consider the problem

$$\min_{c\in\mathbb{R}} \|f - cg\|^2$$

Suppose that

$$\frac{\partial}{\partial c}\|f - cg\|^2 = 0$$

is satisfed at $c = \beta$. Define $\hat{f} = \beta g$ and show that

$$\langle f - \hat{f}, g\rangle = 0$$

and

$$\|f - cg\|^2 = \|f - \hat{f}\|^2 + (c - \beta)^2\|g\|^2$$

Conclude that $c = \beta$ is a solution to the minimization problem.

Exercise 1 (continued)

(2) Let H_1 be a linear subspace of H. Given $f \in H$, consider the problem,

$$\min_{h \in H_1} \|f - h\|^2$$

Suppose $\hat{f} \in H_1$ satisfies

$$\langle f - \hat{f}, h \rangle = 0 \quad \forall h \in H_1.$$

Show that $\hat{f}$ is the unique solution to the minimization problem.

Exercise 1 (continued)

(3) Let H_1 be a subspace of H. Given $f, g \in H$, suppose that $\hat{f}, \hat{g} \in H_1$ satisfy

$$\langle f - \hat{f}, h \rangle = 0 \quad \text{and} \quad \langle g - \hat{g}, h \rangle = 0$$

for all $h \in H_1$. Given $c_1, c_2 \in \mathbb{R}$, consider

$$\min_{h \in H_1} \|c_1 f + c_2 g - h\|^2.$$

What is the solution h^*? Why?

Outline

Minimum-mean-square-error

(X, Y) is a random vector with some joint distribution.

Given value of X, we want to predict Y. Start with simple linear predictor.

$$\hat{Y} = \beta_0 + \beta_1 X$$

Determine β_0, β_1 by solving

$$\min_{\beta_0, \beta_1} E[(Y - \hat{Y})^2].$$

Going to set-up a general tool to solve problems of this form.

Minimum-norm

Use the projection theorem in a vector space with an inner product to solve this problem.

The Vector space: all functions of (X, Y).

The inner product:

$$\langle X, Y \rangle = E[XY].$$

The norm:

$$\|Y\| = \langle Y, Y \rangle^{1/2}.$$

Minimum-mean-square-error problem written as a minimum-norm problem:

$$\min_{\beta_0, \beta_1} \|Y - \hat{Y}\|^2$$

Aside: Why $\langle X, Y \rangle = E[XY]$?

Instead of $\mathrm{Cov}(X, Y)$?

Recall: $\langle X, X \rangle \geq 0$ and equal to zero if and only if $X = 0$.

- Covariance doesn't satisfy this.
- If we used covariance, we are defining "equality" to be "equality up to a constant."

No constant case

First, suppose

$$\hat{Y} = \beta X.$$

We wish to find β that minimizes

$$E[(Y - \beta X)^2] = \|Y - \beta X\|^2.$$

Use the projection theorem! The prediction error will be orthogonal to X.

$$\langle Y - \hat{Y}, X \rangle = 0$$

Let $\mathcal{X}$ be the space of linear functions of X with no constant.

We are looking for the element $\beta^* X \in \mathcal{X}$ that is closest to Y.

No constant case

We have

$$\langle Y - \hat{Y}, X \rangle = \langle Y, X \rangle - \beta \langle X, X \rangle = 0.$$

So,

$$\beta = \frac{\langle Y, X \rangle}{\langle X, X \rangle}$$

or

$$\beta = \frac{E[YX]}{E[X^2]}.$$

Denote the predicted value of Y at the solution as

$$E^*[Y|X] = \beta X.$$

Constant and slope case

Now, suppose

$$\hat{Y} = \beta_0 + \beta_1 X.$$

Find β_0, β_1 that minimizes

$$E[(Y - \beta_0 - \beta_1 X)^2] = \|Y - \beta_0 - \beta_1 X\|^2.$$

Use the projection theorem! The prediction error is orthogonal to $1, X$.

$$\langle Y - \hat{Y}, 1 \rangle = 0, \langle Y - \hat{Y}, X \rangle = 0$$

> Let $\mathcal{X}$ be the space of all linear functions of X - spanned by X and a constant.

Constant and slope

Orthogonality conditions give two equations for two unknowns. Solve them to get

$$\beta_0 = \langle Y, 1\rangle - \beta_1 \langle X, 1\rangle$$

$$\beta_1 = \frac{\langle X, Y\rangle - \langle Y, 1\rangle\langle X, 1\rangle}{\langle X, X\rangle - \langle X, 1\rangle\langle X, 1\rangle}$$

or

$$\beta_0 = E[Y] - \beta_1 E[X]$$

$$\beta_1 = \frac{E[XY] - E[X]E[Y]}{E[X^2] - E[X]^2} = \frac{\mathrm{Cov}(X, Y)}{V(X)}$$

Denote predicted value of Y at solution as

$$E^*[Y|1, X] = \beta_0 + \beta_1 X.$$

Exercise

Let

$$E^*[Y|1, X_1, \ldots, X_K] = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K$$

be the population linear predictor of Y given $1, X_1, \ldots, X_K$. Write this as

$$E^*[Y|X] = X'\beta$$

where $X = (1, X_1, \ldots, X_K)'$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_K)'$. Define

$$U = Y - X'\beta.$$

Show that

$$V(Y) = V(X'\beta) + V(U)$$

Outline

Least squares fit

This orthogonal projection approach is very general.

We'll now switch gears and look at how to apply it to data.

Least squares fit

Let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad x_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

The i-th fitted value is

$$\hat{y}_i = b_0 + b_1 x_i$$

and we want to find the b_0, b_1 to solve

$$\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2.$$

Least squares fit

We'll set-up a minimum norm problem. Define the inner product

$$\langle y, x \rangle = \frac{1}{n} \sum_{i=1}^{n} y_i x_i.$$

The minimum-norm problem is

$$\min_{b_0, b_1} \|y - b_0 x_0 - b_1 x_1\|^2$$

and the solution is obtained via the orthogonal projection of y onto the space spanned by (x_0, x_1).

We have the orthogonality conditions

$$\langle y - \hat{y}, x_0 \rangle = 0$$

$$\langle y - \hat{y}, x_1 \rangle = 0.$$

Least squares fit

The solution is exactly as before

$$b_0 = \langle y, x_0 \rangle - b_1 \langle y, x_1 \rangle$$

$$b_1 = \frac{\langle y, x_1 \rangle - \langle y, x_0 \rangle \langle x, x_0 \rangle}{\langle x_1, x_1 \rangle - \langle x_1, x_0 \rangle \langle x, x_0 \rangle}$$

and plugging in the definition of this norm,

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\bar{yx} - \bar{y}\bar{x}}{\bar{x^2} - \bar{x}^2}$$

where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \bar{xy} = \frac{1}{n} \sum_{i=1}^{n} y_i x_i, \quad \bar{x^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

.

Outline

Omitted variables bias formula

We can use this orthogonal projection tool to derive the typical omitted variables bias formula.

Let (Y, X_1, X_2) be a random vector with some joint distribution. Call

$$E^*[Y|1, X_1, X_2] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

the **long regression**. Call

$$E^*[Y|1, X_1] = \alpha_0 + \alpha_1 X_1$$

the **short regression**. Call

$$E^*[X_2|1, X_1] = \gamma_0 + \gamma_1 X_1$$

the **auxiliary regression**.

Omitted variables bias formula

Can we relate α_1 to β_1? Yes!

Let $U = Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2$. Then,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

with $U \perp 1, X_1, X_2$.

Therefore,

$$\begin{aligned} E^*[Y|1, X_1] &= \beta_0 + \beta_1 X_1 + \beta_2 E^*[X_2|1, X_1] + E^*[U|1, X_1] \\ &= \beta_0 + \beta_1 X_1 + \beta_2(\gamma_0 + \gamma_1 X_1) + 0 \end{aligned}$$

and we re-arrange to get

$$E^*[Y|1, X_1] = (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)X_1.$$

Omitted variable bias formula

Omitted variable bias formula:

$$\alpha_1 = \beta_1 + \beta_2 \gamma_1.$$

There is an exactly identical formula for the least-squares problem.

Exercise

Consider the following model for measurement error:

$$E^*[Y_i|1, \tilde{Z}_i, Z_{i1}, Z_{i2}] = \beta_0 + \beta_1 \tilde{Z}_i$$

$$E^*[Z_{i1}|1, \tilde{Z}_i] = \tilde{Z}_i$$
$$E^*[Z_{i2}|1, \tilde{Z}_i] = \tilde{Z}_i$$

where Z_{i1}, Z_{i2} are noisy measurements of $\tilde{Z}_i$. The population model is expressed in terms of vectors of random variables

$$D_i = (Y_i, \tilde{Z}_i, Z_{i1}, Z_{i2})$$

and we assume the D_i are i.i.d. from some unknown distribution. We observe

$$W_i = (Y_i, Z_{i1}, Z_{i2})$$

for $i = 1, \ldots, n$.

Exercise (continued)

(1) Work out the covariance matrix of (Y_i, Z_{i1}, Z_{i2}) as a function of β_1, $V(\tilde{Z}_i)$ and some additional parameters that you will need to define.

Hint: Define the prediction errors associated with the best linear predictors

$$Y_i = \beta_0 + \beta_1 \tilde{Z}_i + U_i$$

$$Z_{i1} = \tilde{Z}_i + V_{i1}$$

$$Z_{i2} = \tilde{Z}_i + V_{i2}$$

Show these errors are uncorrelated.

(2) Express β_1 as a function of the elements of the covariance matrix in (1).

(3) Consider the linear predictor

$$E^*[Y_i | 1, Z_{i1}] = \pi_0 + \pi_1 Z_{i1}.$$

Is $\pi_1 = \beta_1$, $\pi_1 > \beta_1$ or $\pi_1 < \beta_1$?