Econ 2120, Section 1 - Linear Models

Yibo Sun and Alice Wu

09/01/2021

Welcome!

- Section Times:
  - Mon 2pm-3pm ET, Alice Wu
  - Thurs 3pm-4pm ET, Yibo Sun
- We hold weekly office hours on Zoom, TBD
- Problem sets are due Wednesdays at midnight. You are encouraged to submit in groups of up to 4.

Outline

Linear Algebra (Reference: Friedberg)
- Vector Space
- Inner Product, Norm

Orthogonal Projections
- Orthogonality
- Projection Theorem

Best Linear Predictor
- Minimization Problems
- $E^*[Y|X] = \beta_1 X$
- $E^*[Y|X] = \beta_0 + \beta_1 X$
- Extras: Anscombe's Quartet

Outline

- Linear Algebra (Reference: Friedberg)
  - Vector Space
  - Inner Product, Norm
- Orthogonal Projections
  - Orthogonality
  - Projection Theorem
- Best Linear Predictor
  - Minimization Problems
  - $E^*[Y|X] = \beta_1 X$
  - $E^*[Y|X] = \beta_0 + \beta_1 X$
  - Extras: Anscombe's Quartet

Vector Space

Definition: A vector space V over a field F is a set of elements closed under two operations: addition and scalar multiplication

- $\forall x, y \in V : (x + y) \in V$
- $\forall x \in V, \forall a \in F :\ ax \in V$

such that the following conditions hold

- $x + y = y + x$
- $(x + y) + z = x + (y + z)$
- $\exists\, 0 \in V$ s.t. $x + 0 = x$
- $\exists\, y \in V$ s.t. $x + y = 0$
- $1x = x$
- $(ab)x = a(bx)$
- a(x+y)=ax+ay
- $(a + b)x = ax + bx$

Inner Product $V \times V \to F$

<u>Definition</u>: Let V be a vector space over field F. An inner product on V is a function $V \times V \to F$ that assigns every *ordered* pair $x, y \in V$ a scalar $\langle x, y \rangle \in F$ such that $\forall x, y, z \in V$, $\forall c \in F$:

- Linearity: $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and $\langle ax, y \rangle = c\langle x, y \rangle$
- Symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- Positive Definiteness: $\langle x, x \rangle > 0$ if $x \neq 0$

<u>Notes</u>:
The inner product is a Hermitian form. Inner product space $(V, \langle \cdot, \cdot \rangle)$ is also called a pre-Hilbert / Hermitian space.

Hilbert space = a *complete* inner product space
$(\langle lim_n x_n, lim_{n'} y_{n'} \rangle = lim_{n,n'} \langle x_n, y_{n'} \rangle$

Normed Space

A norm $\|\cdot\| : V \to \mathbb{R}$ such that

- Nonnegativity: $\|x\| \geq 0 \ \forall \ x \in V$, and $\|x\| = 0$ iff $x = 0$
- Multiplication: $\|ax\| = |a|\|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

Given any inner product space $(V, \langle\cdot\rangle)$, the function $\|\cdot\| : V \to \mathbb{R}$ defined by $\|v\| = \sqrt{\langle v, v\rangle}$ is a norm. That is, the inner product *induces* a norm.

Cauchy-Schwarz Inequality

Theorem: $\forall x, y \in V$,

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

Proof: Consider $c \in F = \mathbb{R}$

$$\langle x - cy, x - cy\rangle = \langle x, x\rangle - 2c\langle x, y\rangle + c^2\langle y, y\rangle \geq 0$$

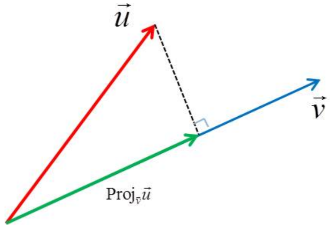
$$\text{let } c = \frac{\langle x, y\rangle}{\langle y, y\rangle}$$

$$\rightarrow \langle x, y\rangle^2 \leq \langle x, x\rangle\langle y, y\rangle$$

Outline

Orthogonality

- Given an inner product space $(V, \langle \cdot, \cdot \rangle)$, x, y are orthogonal $(x \perp y)$ if
$$\langle x, y \rangle = 0$$
- Orthog projection of y onto a vector $x \in V$ is defined by
$$\pi_x(y) = \frac{\langle y, x \rangle}{\|x\|^2} x$$
verify $\langle y - \pi_x(y), x \rangle = 0$



Orthogonality

- Orthog projection of $x \in V$ onto a subspace $W \subset V$:

$$\pi_W(y) = \sum_i \frac{\langle y, w_i \rangle}{\|w_i\|^2} w_i$$

where $\{w_i\}$ is an orthogonal basis for W

  - this projection exists for any y and is also unique
  - $\forall W \subset V : V = W \bigoplus W^{\perp}$

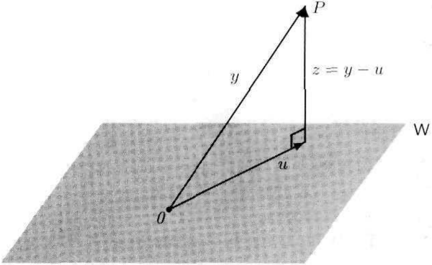


Figure 6.2

Projection Theorem

Theorem

Let V be a vector space (complete with inner product), and $W \subset V$ be a closed subspace. Given any y,

$$w^* = argmin_{w \in W} \|y - w\|$$

iff

$$\langle y - w^*, w \rangle = 0, \, \forall w \in W$$

Note here $w^* = \pi_W(y)$. The distance between y and W is minimized when $(y - w^*) \perp W$

Projection Theorem Proof

$\Longleftarrow$

If $\langle y - w^*, w \rangle = 0$, $\forall w \in W$ (orthogonality condition),

$\forall w \in W$:

$$
\begin{aligned}
\|y - w\|^2 &= \|y - w^* + w^* - w\|^2 \\
&= \langle y - w^*, y - w^* \rangle + 2\langle y - w^*, w \rangle + \langle w^* - w, w^* - w \rangle \\
&= \|y - w^*\|^2 + 0 + \|w^* - w\|^2 \geq \|y - w^*\|^2
\end{aligned}
$$

$\Longrightarrow$

Given $w^* = argmin_{w \in W} \|y - w\|$, suppose
$\exists w \in W$: $\langle y - w^*, w \rangle = t > 0$.
WLOG, $\|w\| = 1$, and let $w' = w^* + tw \in W$, we can show:

$$
\begin{aligned}
\|y - w'\|^2 &= \|y - (w^* + tw)\|^2 \\
&= \|y - w^*\|^2 - 2t\langle y - w^*, w \rangle + t^2\|w\|^2 \\
&< \|y - w^*\|^2 \ \Rightarrow\Leftarrow
\end{aligned}
$$

More on Projection ThM

- w^* exists for any $y \in V$ and is unique (proof by contraditction)
- We can find w^* by a system of orthogonality conditions (linear equations)
$$(y - w^*) \perp w_i$$
given any basis w_i for W

Outline

Linear Algebra (Reference: Friedberg)

- Vector Space
- Inner Product, Norm

Orthogonal Projections

- Orthogonality
- Projection Theorem

Best Linear Predictor

- Minimization Problems
- $E^*[Y|X] = \beta_1 X$
- $E^*[Y|X] = \beta_0 + \beta_1 X$
- Extras: Anscombe's Quartet

Predictors

Given random variables (Y, X) in a vector space V, can we predict Y given X?

- next time we will focus on optimal predictor

$$E[Y|X] = argmin_f E[(Y - f(X))^2]$$

- today, start with linear predictors, and we call the best linear predictor

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

Minimization Problems

1) Minimum-mean-square-error Predictors:

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

where

$$\beta = argmin_b E[(Y - Xb)^2]$$

- FOC: $-2E[X'(Y - X\beta)] = 0$
- $\beta = E[X'X]^{-1}E[X'Y]$ uniquely determined if $E[X'X]$ is nonsingular

Minimization Problems

2) Minimum-norm Predictors: Consider an inner product $\langle Y, X \rangle = E[YX]$ and the associated norm $\|Y\| = \langle Y, Y \rangle^{1/2}$.

Rewrite the BLP as

$$E^*[Y|X] = X\beta = argmin_{Xb \in span(X)} \|Y - Xb\|^2$$

- Can we define $\langle Y, X \rangle = cov(Y, X)$?
- We are looking for $\pi_X(Y)$, the linear projection of Y onto the subspace spanned by X.
- Solve orthogonality conditions: $Y - X\beta \perp span(X)$ or

$$\langle Y - X\beta, x \rangle = 0$$

for any $x \in span(X)$

No-constant Case

$dim(X) = 1$, look for the best linear predictor without constant $\hat{Y} = \beta_1 X$ where $\beta_1 = argmin_{b_1} \| Y - b_1 X \|$
By Projection ThM, sufficient to solve orthogonality condition for β_1:

$$\langle Y - \beta_1 X, X \rangle = 0$$
$$\rightarrow \beta_1 = \frac{\langle Y, X \rangle}{\langle X, X \rangle} = \frac{E[YX]}{E[X^2]}$$

Sample analog: define

$$\langle y, x \rangle = \frac{1}{n} \sum_i y_i x_i$$

Orthog condition $\langle y - b_1 x, x \rangle = 0$ yields $\hat{\beta}_1 = \frac{1/n \sum_i y_i x_i}{1/n \sum_i x_i^2}$

Adding a Constant

Look for the best linear predictor with constant $\hat{Y} = \beta_0 + \beta_1 X$ where $(\beta_0, \beta_1) = argmin_b \| Y - b_0 - b_1 X \|$

Now we want to find the orthog projection of Y onto the subspace spanned by $\{1, X\}$ (as a basis as long as X not constant).

Orthogonality conditions:

- $\langle Y - \beta_0 - \beta_1 X, 1 \rangle = 0$
- $\langle Y - \beta_0 - \beta_1 X, X \rangle = 0$

$$\beta_0 = E[Y] - \beta_1 E[X]$$
$$\beta_1 = \frac{E[(Y - E[Y])X]}{E[(X - E[X])X]} = \frac{cov(Y, X)}{var(X)}$$

Adding a Constant

If you would like to use more linear algebra,

- what is an orthogonal basis for the subspace $span(1, X)$?
  - note $\pi_1(X) = E[X]$
  - by Gram-Schmidt, $\{1, X - E[X]\}$ or $\{1\}$ if X is a constant
- orthog projection of Y onto $W = span(1, X)$ can be written as:

$$\pi_W(Y) = \sum_i \frac{\langle Y, w_i \rangle}{\|w_i\|^2} w_i$$

$$= E[Y] + \frac{E[Y(X - E[X])]}{E[(X - E[X])^2]}(X - E[X])$$

Adding a Constant

- By Projection ThM, the orthogonal projection exists and is unique.
$$E^*[Y|X] = argmin_{w \in span(X)} \|Y - w\|^2$$

- if X is also a constant (collinear with 1), we have the unique orthog projection $E[Y]$ as the best linear predictor.

- We won't have a unique representation $\hat{Y} = \beta_0 + \beta_1 X$ (can't solve FOC in min-MSE)

$E^*[Y|X]$ vs. $E[Y|X]$

- Econometrics problems can often be thought of as prediction problems. The Best Linear Predictor uses information about X to predict Y.
  - X *contains information* about Y if knowing it improves our prediction
- $E^*[Y|1, X]$ characterizes some of the informational content of X. It does not tell the whole story.
- Canonical example: Anscombe's quartet

Anscombe's Quartet (1973)