CS70

Conditional Expectation and Estimations

Kelvin Lee

UC Berkeley

April 20, 2021

Overview

1. LLSE
   - Projection
2. MMSE
   - Orthogonality Property of MMSE
3. Properties of Conditional Expectation

LLSE

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

Theorem

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

Theorem

$$L[Y|X] = a + bX = \mathbb{E}[Y] + \frac{Cov(X,Y)}{Var(X)}(X - \mathbb{E}[X]).$$

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

Theorem

$$L[Y|X] = a + bX = \mathbb{E}[Y] + \frac{Cov(X, Y)}{Var(X)}(X - \mathbb{E}[X]).$$

Proof:

LLSE

The Linear Least Squares Estimate (*LLSE*) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

Theorem

$$L[Y|X] = a + bX = \mathbb{E}[Y] + \frac{Cov(X, Y)}{Var(X)}(X - \mathbb{E}[X]).$$

Proof:

Expand $\mathbb{E}[(Y - (a + bX))^2]$ and optimize over a and b.

LLSE

The Linear Least Squares Estimate ($LLSE$) of Y given X, denoted by $L[Y|X]$ is the linear function $a + bX$ that minimizes

$$\mathbb{E}[(Y - (a + bX))^2].$$

Theorem

$$L[Y|X] = a + bX = \mathbb{E}[Y] + \frac{Cov(X, Y)}{Var(X)}(X - \mathbb{E}[X]).$$

Proof:

Expand $\mathbb{E}[(Y - (a + bX))^2]$ and optimize over a and b. Plug in optimal a, b to $a + bX$. □

Projection

Projection

$L[Y|X]$ can be interpreted as the **projection** of Y onto the set of linear functions of X, call it $\mathcal{L}(X)$.

Projection

$L[Y|X]$ can be interpreted as the **projection** of Y onto the set of linear functions of X, call it $\mathcal{L}(X)$.

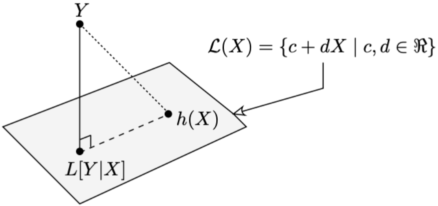

Figure: $L[Y|X]$ is the projection of Y onto $\mathcal{L}(X)$.

Orthogonality

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.
- Thus, $L[Y|X] = a + bX$ is the projection of Y onto $\mathcal{L}(X)$ if $Y - L[Y|X]$ is orthogonal to every linear function of X, i.e., if

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.
- Thus, $L[Y|X] = a + bX$ is the projection of Y onto $\mathcal{L}(X)$ if $Y - L[Y|X]$ is orthogonal to every linear function of X, i.e., if

$$\forall c, d \in \mathbb{R}, \quad \mathbb{E}[(Y - (a + bX))(c + dX)] = 0.$$

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.
- Thus, $L[Y|X] = a + bX$ is the projection of Y onto $\mathcal{L}(X)$ if $Y - L[Y|X]$ is orthogonal to every linear function of X, i.e., if

$$\forall c, d \in \mathbb{R}, \quad \mathbb{E}[(Y - (a + bX))(c + dX)] = 0.$$

Equivalently,

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.
- Thus, $L[Y|X] = a + bX$ is the projection of Y onto $\mathcal{L}(X)$ if $Y - L[Y|X]$ is orthogonal to every linear function of X, i.e., if

$$\forall c, d \in \mathbb{R}, \quad \mathbb{E}[(Y - (a + bX))(c + dX)] = 0.$$

Equivalently,

$$\mathbb{E}[Y] = a + b\mathbb{E}[X] \text{ and } \mathbb{E}[(Y - (a + bX))X] = 0.$$

Orthogonality

- Two vectors V and W are **orthogonal** to each other if and only if $\mathbb{E}[VW] = 0$.
- Thus, $L[Y|X] = a + bX$ is the projection of Y onto $\mathcal{L}(X)$ if $Y - L[Y|X]$ is orthogonal to every linear function of X, i.e., if

$$\forall c, d \in \mathbb{R}, \quad \mathbb{E}[(Y - (a + bX))(c + dX)] = 0.$$

Equivalently,

$$\mathbb{E}[Y] = a + b\mathbb{E}[X] \text{ and } \mathbb{E}[(Y - (a + bX))X] = 0.$$

This is known as the **projection property**.

MMSE

MMSE

- The *MMSE* of Y given X is given by $\mathbb{E}[Y|X]$, which is the conditional expectation of Y given X.

MMSE

- The *MMSE* of Y given X is given by $\mathbb{E}[Y|X]$, which is the conditional expectation of Y given X.
- The **conditional expectation** of Y given X is defined by

MMSE

- The *MMSE* of Y given X is given by $\mathbb{E}[Y|X]$, which is the conditional expectation of Y given X.
- The **conditional expectation** of Y given X is defined by

$$\mathbb{E}[Y|X=x] = \sum_y y \cdot \mathbb{P}(Y=y|X=x) = \sum_y y \cdot \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x)}.$$

Orthogonality Property of *MMSE*

Orthogonality Property of *MMSE*

Lemma (Orthogonality Property of *MMSE*)

For any function $\phi(\cdot)$*, one has*

Orthogonality Property of *MMSE*

Lemma (Orthogonality Property of *MMSE*)

For any function $\phi(\cdot)$, one has

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0.$$

Orthogonality Property of *MMSE*

Lemma (Orthogonality Property of *MMSE*)

For any function $\phi(\cdot)$, *one has*

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0.$$

Moreover, if the function $g(X)$ *is such that*

Orthogonality Property of *MMSE*

Lemma (Orthogonality Property of *MMSE*)

For any function $\phi(\cdot)$, one has

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0.$$

Moreover, if the function $g(X)$ is such that

$$\forall \phi(\cdot), \quad \mathbb{E}[(Y - g(X))\phi(X)] = 0,$$

Orthogonality Property of *MMSE*

Lemma (Orthogonality Property of *MMSE*)

For any function $\phi(\cdot)$*, one has*

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0.$$

Moreover, if the function $g(X)$ *is such that*

$$\forall \phi(\cdot), \quad \mathbb{E}[(Y - g(X))\phi(X)] = 0,$$

then $g(X) = \mathbb{E}[Y|X]$*.*

Properties of Conditional Expectation

Properties of Conditional Expectation

- Linearity:

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1|X] + a_2 \mathbb{E}[Y_2|X];$$

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1 | X] + a_2 \mathbb{E}[Y_2 | X];$$

- Factoring known values:

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1 | X] + a_2 \mathbb{E}[Y_2 | X];$$

- Factoring known values:

$$\mathbb{E}[h(X) Y \mid X] = h(X) \mathbb{E}[Y | X];$$

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1|X] + a_2 \mathbb{E}[Y_2|X];$$

- Factoring known values:

$$\mathbb{E}[h(X) Y \mid X] = h(X)\mathbb{E}[Y|X];$$

- Law of iterated expectation:

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1 | X] + a_2 \mathbb{E}[Y_2 | X];$$

- Factoring known values:

$$\mathbb{E}[h(X) Y \mid X] = h(X) \mathbb{E}[Y | X];$$

- Law of iterated expectation:

$$\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[Y].$$

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1|X] + a_2 \mathbb{E}[Y_2|X];$$

- Factoring known values:

$$\mathbb{E}[h(X)Y \mid X] = h(X)\mathbb{E}[Y|X];$$

- Law of iterated expectation:

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

- Independence:

Properties of Conditional Expectation

- Linearity:

$$\mathbb{E}[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 \mathbb{E}[Y_1 | X] + a_2 \mathbb{E}[Y_2 | X];$$

- Factoring known values:

$$\mathbb{E}[h(X) Y \mid X] = h(X) \mathbb{E}[Y | X];$$

- Law of iterated expectation:

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

- Independence:

$$\mathbb{E}[Y|X] = \mathbb{E}[Y].$$

Problem Time!