# I. EXERCISE 3: MEAN VALUES AND VARIANCES IN LINEAR REGRESSION

Assuming the existence of a function $f(\boldsymbol{x})$ as well as a normally distributed error $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ who describe our data.

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\varepsilon} \tag{1}$$

We approximate the function with Linear regression, OLS. Here f is approximated by $\tilde{\boldsymbol{y}}$. We minimize $(\boldsymbol{y} - \tilde{\boldsymbol{y}})^2$, with

$$\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} \tag{2}$$

The $\boldsymbol{X}$ here is the design- or feature-matrix.

## A.

show the expectation value of $\boldsymbol{y}$ for a given element $i$ is

$$\mathbb{E}(y_i) = \boldsymbol{X}_{i,*}\boldsymbol{\beta}, \tag{3}$$

and the variance is

$$\text{Var}(y_i) = \sigma^2. \tag{4}$$

### *1.*

the data set $\boldsymbol{y}$ is assumed modelled as a sum of the deterministic $f(\boldsymbol{x})$ and the stochastic noise $\varepsilon$. The mean of the set then should be modelled as

$$\langle \boldsymbol{y}_i \rangle = \frac{1}{n}\sum_{i=0}^{n-1}\left(f(\boldsymbol{x_i}) + \varepsilon_i\right) = \frac{1}{n}\Big[\sum_{i=0}^{n-1} f(\boldsymbol{x}_i)\ \sum_{i=0}^{n-1}\varepsilon_i\Big] \tag{5}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} f(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=0}^{n-1}\boldsymbol{X}_{i,*}\boldsymbol{\beta} = \boldsymbol{X}_{i,*}\boldsymbol{\beta} \tag{6}$$

Where the $\boldsymbol{X}$ matrix and the $\boldsymbol{\beta}$ vector are both deterministic, And so the mean is the values themselves.

As for the var$(\boldsymbol{y}_i)$

$$\text{Var}(\boldsymbol{y}_i) = \left\langle (\boldsymbol{y}_i - \langle \boldsymbol{y}_i \rangle)^2 \right\rangle = \left\langle \boldsymbol{y}_i^2 \right\rangle - \langle \boldsymbol{y}_i \rangle^2 \tag{7}$$

$$= \left\langle (\boldsymbol{X}_{i,*}\boldsymbol{\beta} + \varepsilon)^2 \right\rangle - (\boldsymbol{X}_{i,*}\boldsymbol{\beta})^2 \tag{8}$$

$$= (\boldsymbol{X}_{i,*}\boldsymbol{\beta})^2 + 2\langle \varepsilon \rangle\, \boldsymbol{X}_{i,*}\boldsymbol{\beta} \left\langle \varepsilon^2 \right\rangle - (\boldsymbol{X}_{i,*}\boldsymbol{\beta})^2 \tag{9}$$

$$= \left\langle \varepsilon^2 \right\rangle = \text{Var}(\varepsilon) = \sigma^2 \tag{10}$$

Q.E.D

## B.

With the OLS expression for the parameters $\boldsymbol{\beta}$, show

$$\langle \boldsymbol{\beta} \rangle = \boldsymbol{\beta} \tag{11}$$

*1.*

The OLS expression for $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T Y}. \tag{12}$$

$$\tag{13}$$

So the mean value for beta is,

$$\langle \boldsymbol{\beta} \rangle = \left\langle \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T Y} \right\rangle = \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T} \langle \boldsymbol{Y} \rangle \tag{14}$$

$$= \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T X} \boldsymbol{\beta} = \boldsymbol{\beta} \tag{15}$$

**C.**

Show that the variance of $\boldsymbol{\beta}$ is

$$\mathrm{Var}(\boldsymbol{\beta}) = \sigma^2 (\boldsymbol{X^T X})^{-1}. \tag{16}$$

We'll start with the definition of variance,

$$\mathrm{Var}(\boldsymbol{\beta}) = \left\langle (\boldsymbol{beta} - \langle \boldsymbol{beta} \rangle)^2 \right\rangle = \left\langle (\boldsymbol{beta} - \langle \boldsymbol{beta} \rangle)(\boldsymbol{beta} - \langle \boldsymbol{beta} \rangle)^T \right\rangle \tag{17}$$

$$= \left\langle \left( \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T Y} - \boldsymbol{\beta} \right) \left( \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T Y} - \boldsymbol{\beta} \right)^T \right\rangle \tag{18}$$

$$= \left\langle (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y Y^T X} (\boldsymbol{X^T X})^{-1} - (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y} \boldsymbol{\beta^T} - \boldsymbol{\beta Y^T X} (\boldsymbol{X^T X})^{-1} + \boldsymbol{\beta\beta^T} \right\rangle \tag{19}$$

$$= (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \left\langle \boldsymbol{Y Y^T} \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} \tag{20}$$

$$- (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \langle \boldsymbol{Y} \rangle \boldsymbol{\beta^T} \tag{21}$$

$$- \boldsymbol{\beta} \left\langle \boldsymbol{Y^T} \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} \tag{22}$$

$$+ \boldsymbol{\beta\beta^T} \tag{23}$$

$$= (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \left\langle \boldsymbol{Y Y^T} \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} \tag{24}$$

$$- (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T X} \boldsymbol{\beta\beta^T} \tag{25}$$

$$- \boldsymbol{\beta\beta^T} \boldsymbol{X^T X} (\boldsymbol{X^T X})^{-1} \tag{26}$$

$$+ \boldsymbol{\beta\beta^T} \tag{27}$$

The products of the matrices and their inverse canceling to identities

$$\mathrm{Var}(\boldsymbol{\beta}) = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \left\langle \boldsymbol{Y Y^T} \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} - \boldsymbol{\beta\beta^T} \tag{28}$$

$$= (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \left\langle (\boldsymbol{X\beta} + \boldsymbol{\varepsilon})(\boldsymbol{\beta^T X^T} + \boldsymbol{\varepsilon}) \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} - \boldsymbol{\beta\beta^T} \tag{29}$$

$$= (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T} \left\langle \boldsymbol{X\beta\beta^T X^T} + \boldsymbol{X\beta\varepsilon} + \boldsymbol{\varepsilon\beta^T X^T} + \varepsilon^2 \right\rangle \boldsymbol{X} (\boldsymbol{X^T X})^{-1} - \boldsymbol{\beta\beta^T} \tag{30}$$

$$= \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T} \left( \boldsymbol{X\beta\beta^T X^T} + \left\langle \varepsilon^2 \right\rangle \right) \boldsymbol{X} (\boldsymbol{X^T X})^{-1} - \boldsymbol{\beta\beta^T} \tag{31}$$

Here, we have that $\left\langle \varepsilon^2 \right\rangle = \sigma^2$. Also,

$$\left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T X \beta\beta^T X^T X} \left(\boldsymbol{X^T X}\right)^{-1} = \mathbb{1}\boldsymbol{\beta\beta^T}\mathbb{1} = \boldsymbol{\beta\beta^T}$$

This lets us reduce the expression

$$\mathrm{Var}(\boldsymbol{\beta}) = \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T} \sigma^2 \boldsymbol{X} \left(\boldsymbol{X^T X}\right)^{-1} \tag{32}$$

$$= \left(\boldsymbol{X^T X}\right)^{-1} \boldsymbol{X^T X} \sigma^2 \left(\boldsymbol{X^T X}\right)^{-1} \tag{33}$$

$$= \mathbb{1}\sigma^2 \left(\boldsymbol{X^T X}\right)^{-1} \tag{34}$$

Q.E.D