Linear regression

Machine Learning 2021-2022 - UMONS
Souhaib Ben Taieb

1

Consider the following optimization problem:

$$(\beta_0^*, \beta_1^*) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{\operatorname{argmin}} \; E_{\text{out}}(\beta_0, \beta_1) := \mathbb{E}_{x,y}[(y-(\beta_0+\beta_1 x))^2]$$

where $x, y \in \mathbb{R}$.

Show that the solution is given by

$$\begin{aligned}
\beta_1^* &= \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)}, \\
\beta_0^* &= \mathbb{E}[y] - \beta_1 \mathbb{E}[x].
\end{aligned}$$

We have

$$\begin{aligned}
E_{\text{out}}(\beta_0, \beta_1) &= \mathbb{E}[(y-(\beta_0+\beta_1 x))^2] \\
&= \mathbb{E}[y^2] - 2\beta_0\mathbb{E}[y] - 2\beta_1\mathbb{E}[xy] + \mathbb{E}[(\beta_0+\beta_1 x)^2] \\
&= \mathbb{E}[y^2] - 2\beta_0\mathbb{E}[y] - 2\beta_1(\operatorname{Cov}(x,y) + \mathbb{E}[x]\mathbb{E}[y]) \\
&+ \mathbb{E}[(\beta_0+\beta_1 x)^2] \\
&= \mathbb{E}[y^2] - 2\beta_0\mathbb{E}[y] - 2\beta_1(\operatorname{Cov}(x,y) + \mathbb{E}[x]\mathbb{E}[y]) \\
&+ \beta_0^2 + \beta_1^2\mathbb{E}[x^2] + 2\beta_0\beta_1\mathbb{E}[x] \\
&= \mathbb{E}[y^2] - 2\beta_0\mathbb{E}[y] - 2\beta_1\operatorname{Cov}(x,y) - 2\beta_1\mathbb{E}[x]\mathbb{E}[y] \\
&+ \beta_0^2 + \beta_1^2\operatorname{Var}(x) + \beta_1^2(\mathbb{E}[x])^2 + 2\beta_0\beta_1\mathbb{E}[x]
\end{aligned}$$

where we used the following identities:

$$\operatorname{Cov}(x,y) = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y], \quad \operatorname{Var}(x) = \mathbb{E}[x^2] - (\mathbb{E}[x])^2.$$

Taking the partial derivatives of $E_{\text{out}}(\beta_0, \beta_1)$ with respect to β_0 and β_1 and equaling to zero yields :

$$\begin{aligned}
\frac{\partial E_{\text{out}}(\beta_0,\beta_1)}{\partial \beta_0} = 0 &\iff -2\mathbb{E}[y] + 2\beta_0 + 2\beta_1\mathbb{E}[x] = 0 \\
&\iff \beta_0 = \mathbb{E}[y] - \beta_1\mathbb{E}[x]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E_{\text{out}}(\beta_0,\beta_1)}{\partial \beta_1} = 0 &\iff -2\operatorname{Cov}(x,y) - 2\mathbb{E}[x]\mathbb{E}[y] + 2\beta_1\operatorname{Var}(x) + 2\beta_1\left(\mathbb{E}[x]\right)^2 + 2\beta_0\mathbb{E}[x] = 0 \\
&\iff -2\operatorname{Cov}(x,y) - 2\mathbb{E}[x]\mathbb{E}[y] + 2\beta_1\operatorname{Var}(x) + 2\beta_1\left(\mathbb{E}[x]\right)^2 + 2\mathbb{E}[x]\mathbb{E}[y] - 2\beta_1\left(\mathbb{E}[x]\right)^2 = 0 \\
&\iff \beta_1 = \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)}
\end{aligned}$$

2

Consider a dataset $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^n$ with $x_i, y_i \in \mathbb{R}$, and the following optimization problem:

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{\operatorname{argmin}} E_{\text{in}}(\beta_0, \beta_1),$$

where

$$E_{\text{in}}(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2.$$

Prove that the minimizing values $\hat{\beta}_1$ and $\hat{\beta}_0$ are given by

$$\begin{aligned}\hat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x},\end{aligned}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

$$\begin{aligned}\frac{\partial \text{RSS}}{\partial \beta_0} = 0 &\iff -2\sum_i (y_i - (\beta_0 + \beta_1 x_i)) = 0, \\ &\iff \sum_i y_i = n\beta_0 + \beta_1 \sum_i x_i \\ &\iff \beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}\end{aligned}$$

$$\begin{aligned}\frac{\partial \text{RSS}}{\partial \beta_1} = 0 &\iff -2\sum_i x_i(y_i - (\beta_0 + \beta_1 x_i)) = 0 \\ &\iff \sum_i x_i y_i = \sum_i (\beta_0 x_i + \beta_1 x_i^2)) \\ &\iff \sum_i x_i y_i = n\beta_0 \bar{x} + \beta_1 \sum_i x_i^2 \\ &\iff \sum_i x_i y_i = n(\bar{y} - \hat{\beta}_1 \bar{x})\bar{x} + \beta_1 \sum_i x_i^2 \\ &\iff \sum_i x_i y_i - n\bar{x}\bar{y} = \beta_1 (\sum_i x_i^2 - n\bar{x}^2) \\ &\iff \beta_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2}\end{aligned}$$

Using the following equalities proves the result.

$$\begin{aligned}\sum_i (x_i - \bar{x})^2 &= \sum_i x_i^2 - n\bar{x}^2 \\ \sum_i (x_i - \bar{x})(y_i - \bar{y}) &= \sum_i x_i y_i - n\bar{x}\bar{y}\end{aligned}$$

3

We now assume the data has been generated by the following model

$$y_i = f(x_i) + \varepsilon_i,$$

where x_i is fixed (non-random), ε_i are i.i.d. with $E[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$.

Show that the variance of $\hat{\beta}_1$ is given by $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$. You can use the following equalities

$$\begin{aligned}
\sum_i (x_i - \bar{x})(y_i - \bar{y}) &= \sum_i (x_i - \bar{x}) y_i - \sum_i (x_i - \bar{x}) \bar{y} \\
&= \sum_i (x_i - \bar{x}) y_i \\
&= \sum_i (x_i - \bar{x})(\beta_0^* + \beta_1^* x_i + \varepsilon_i)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_1) &= \mathrm{Var}\left(\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}\right) \\
&= \mathrm{Var}\left(\frac{\sum_i (x_i - \bar{x})(\beta_0^* + \beta_1^* x_i + \varepsilon_i)}{\sum_i (x_i - \bar{x})^2}\right), \quad \text{using the previous equalities} \\
&= \mathrm{Var}\left(\frac{\sum_i (x_i - \bar{x})\varepsilon_i}{\sum_i (x_i - \bar{x})^2}\right), \quad \text{only } \varepsilon_i \text{ is random} \\
&= \frac{\sum_i (x_i - \bar{x})^2 \mathrm{Var}(\varepsilon_i)}{\left(\sum_i (x_i - \bar{x})^2\right)^2}, \quad \text{independence of } \varepsilon_i \text{ and, } \mathrm{Var}(kX) = k^2 \mathrm{Var}(X) \\
&= \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}
\end{aligned}$$

Show that the variance of $\hat{\beta}_0$ is given by

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\bar{y} - \hat{\beta}_1 \bar{x}) \\
&= \mathrm{Var}(\bar{y}) + (\bar{x})^2 \mathrm{Var}(\hat{\beta}_1) - 2\mathrm{Cov}(\bar{y}, \hat{\beta}_1 \bar{x}) \\
&= \mathrm{Var}(\bar{y}) + (\bar{x})^2 \mathrm{Var}(\hat{\beta}_1) - 2\bar{x}\mathrm{Cov}(\bar{y}, \hat{\beta}_1)
\end{aligned}$$

where

$$\mathrm{Var}(\bar{y}) = \frac{1}{n^2} \mathrm{Var}(\sum_i^n y_i) = \frac{\sigma^2}{n},$$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$
\begin{aligned}
\mathrm{Cov}(\bar{y}, \hat{\beta}_1) &= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{\sum_{j=1}^{n}(x_j-\bar{x})y_j}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \\
&= \frac{1}{n}\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\mathrm{Cov}\left(\sum_{i=1}^{n} y_i, \sum_{j=1}^{n}(x_j-\bar{x})y_j\right) \\
&= \frac{1}{n}\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})\sum_{j=1}^{n}\mathrm{Cov}(y_i, y_j) \\
&= \frac{1}{n}\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})\sigma^2 \\
&= 0 \quad \left(\text{since } \sum_{i=1}^{n}(x_i-\bar{x}) = 0\right).
\end{aligned}
$$

4

Under the same set of assumptions as the previous exercise, show that the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased, i.e. $\text{Bias}(\hat{\beta}_0) = 0$ and $\text{Bias}(\hat{\beta}_1) = 0$.

$$\text{Bias}(\hat{\beta}_1) = \mathbb{E}[\hat{\beta}_1] - \beta_1^* \tag{1}$$

$$\begin{aligned}
\text{Bias}(\hat{\beta_1}) &= \mathbb{E}[\hat{\beta}_1] - \beta_1^* \\
&= \mathbb{E}\left[\frac{\frac{1}{n}\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2}\right] - \beta_1^* \\
&= \frac{\mathbb{E}\left[\frac{1}{n}\sum_i^n (x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})\right]}{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2} - \beta_1^* \quad x_i\text{'s are non-random.} \\
&= \frac{\mathbb{E}\left[\frac{1}{n}\sum_i^n x_i(\beta_0^* + \beta_1^* x_i + \varepsilon_i) - \bar{x}(\beta_0^* + \beta_1^*\bar{x} + \bar{\varepsilon})\right]}{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2} - \beta_1^* \\
&= \frac{\frac{1}{n}\sum_i^n x_i(\beta_0^* + \beta_1^* x_i + \mathbb{E}[\varepsilon_i]) - \bar{x}(\beta_0^* + \beta_1^*\bar{x} + \mathbb{E}[\bar{\varepsilon}])}{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2} - \beta_1^* \\
&= \frac{\beta_0^*\bar{x} + \beta_1^*\overline{x^2} - \beta_0^*\bar{x} + \beta_1^*(\bar{x})^2}{\frac{1}{n}\sum_i^n \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)} - \beta_1^* \quad \text{as } \mathbb{E}[\varepsilon_i] = \mathbb{E}\left[\frac{1}{n}\sum_i^n \varepsilon_i\right] = \mathbb{E}[\bar{\varepsilon}] = 0 \\
&= \frac{\beta_1^*\left(\overline{x^2} + (\bar{x})^2\right)}{\overline{x^2} + (\bar{x})^2} - \beta_1^* \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Bias}(\hat{\beta}_0) &= \mathbb{E}[\hat{\beta}_0] - \beta_0^* \\
&= \mathbb{E}[\bar{y} - \hat{\beta}_1\bar{x}] - \beta_0^* \\
&= \mathbb{E}[\bar{y}] - \bar{x}\mathbb{E}[\hat{\beta}_1] - \beta_0^* \\
&= \mathbb{E}[\beta_0^* + \beta_1^*\bar{x} + \bar{\varepsilon}] - \beta_1^*\bar{x} - \beta_0^* \\
&= 0
\end{aligned}$$