Simple Linear Regression

© Malay K. Das, 210 Southern Lab, ph-7359, mkdas@iitk.ac.in

Office hours: W 1030-1130, SL-210

Previously:

1. Course introduction
2. kNN

Today:

1. Simple linear regression
2. Error estimation
3. Example

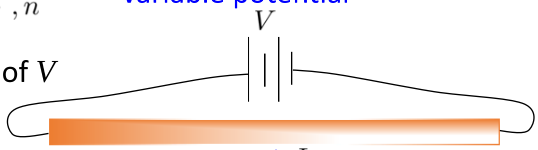
HW due:

August 16, 2024

1

Experimental results: $I_i = \dfrac{V_i}{R} + \epsilon_i \qquad i = 1, 2, \cdots, n$

Zero mean random noise independent of V



Fundamental idea of **Regression**

We try to estimate R in such a way that the **Empirical Risk** is minimized

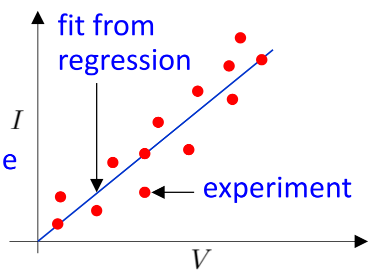
Goal of **Regression**

If mechanistic model (such as $V = IR$) is NOT available

- developing **empirical (data-driven) model**

If mechanistic/empirical model is available

- finding parameter



Finally, **regression estimates** dependent variable (such as $I_\star$) at some point of interest $V_\star$

Regression is a part of Supervised ML

2

Supervised machine learning

supervised, since training labels are known

feature: $\mathbf{x}_i \in \mathbb{R}^n$ label: $y_i \in \mathbb{R}$

Given a set of training data $\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$

we wish to estimate $\widehat{y}_\star(x_\star)$

'hat' sign indicates estimation (NOT exact)

Classification: categorical (integer) labels **Regression:** numerical (real) labels

---

feature space $\mathcal{X}$ $\xrightarrow{h}$ label space $\mathcal{Y}$

learning output:

hypothesis: $h : \mathcal{X} \to \mathcal{Y}$ } inductive learning

impossible without bias

$h \in \mathcal{H}$

created from inductive bias

we cannot test all possible hypotheses, instead we select hypothesis from a hypotheses space $\mathcal{H}$

inductive bias: assumptions generated from constraints, domain knowledge

3

Simple Linear Regression

(x_i, y_i)

For training data $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^n$ estimate $\widehat{y}_\star(x_\star)$

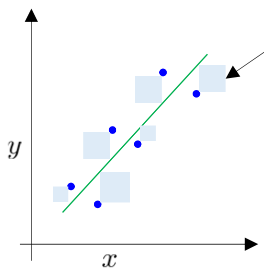
in simple linear regression $x_i \in \mathbb{R}, y_i \in \mathbb{R}$

residual (not the noise from experiments)

we assume a hypothesis $\widehat{y} = w_0 + w_1 x$ $y = \widehat{y} + e$ $E(w_0, w_1) = \dfrac{1}{n}\sum_{i=1}^{n} L_i$

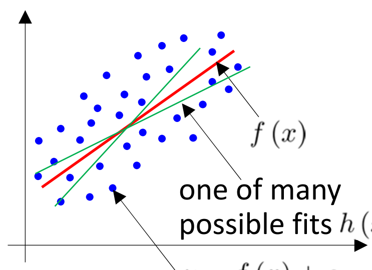
Least square estimate minimizes **Cost Function**

where **Loss Function** $L_i = e_i^2$



Least square estimate minimizes sum of all areas of the squares

Same hypothesis leads to different results depending on (a) training data, (b) definition of cost function



$f(x)$

one of many possible fits $h(x)$

$y = f(x) + \epsilon$

zero-mean random Gaussian noise (computer generated)

Empirical risk (cost function) minimization $h(x) \approx f(x)$

4

Simple Linear Regression: $\widehat{y} = w_0 + w_1 x \quad y = \widehat{y} + e$ Training set: $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^{n}$

To find w_0, w_1 we minimize **Cost Function** $E(w_0, w_1) = \dfrac{1}{n}\sum_{i=1}^{n} e_i^2$

$$E(w_0, w_1) = \frac{1}{n}\left[(y_1 - w_0 - w_1 x_1)^2 + (y_2 - w_0 - w_1 x_2)^2 + \cdots + (y_n - w_0 - w_1 x_n)^2\right]$$

To minimize $E(w_0, w_1)$

we enforce $\nabla E(w_0, w_1) = \mathbf{0} = \begin{bmatrix} \dfrac{\partial E}{\partial w_0} & \dfrac{\partial E}{\partial w_1} \end{bmatrix}^T$

$$\frac{\partial E}{\partial w_0} = 0 \quad \frac{\partial E}{\partial w_1} = 0 \longrightarrow w_0, w_1$$

and verify $\nabla^2 E$ is positive definite

positive definiteness requires all eigenvalues to be positive

checked using Sylvester condition

$$\frac{\partial^2 E}{\partial w_0^2}\frac{\partial^2 E}{\partial w_1^2} - \left(\frac{\partial^2 E}{\partial w_0 \partial w_1}\right)^2 > 0 \quad \frac{\partial^2 E}{\partial w_0^2} > 0$$

$$\nabla^2 E = \begin{bmatrix} \dfrac{\partial^2 E}{\partial w_0^2} & \dfrac{\partial^2 E}{\partial w_0 \partial w_1} \\ \dfrac{\partial^2 E}{\partial w_0 \partial w_1} & \dfrac{\partial^2 E}{\partial w_1^2} \end{bmatrix}$$ (Hessian)

Cost Function $E(w_0, w_1) = \dfrac{1}{n}\Big[(y_1 - w_0 - w_1 x_1)^2 + (y_2 - w_0 - w_1 x_2)^2 + \cdots + (y_n - w_0 - w_1 x_n)^2\Big]$

we enforce

$$\frac{\partial E}{\partial w_0} = 0 \quad \frac{\partial E}{\partial w_1} = 0$$

Discrete data set: $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^{n}$

to estimate $\widehat{y}_\star(x_\star)$

$\widehat{y}_\star \longleftarrow$ estimation of y

$x_\star \longleftarrow$ input

Regression model: $\widehat{y} = w_0 + w_1 x$

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

$$\frac{\partial E}{\partial w_0} = 0 = -\frac{2}{n}\left[(y_1 - w_0 - w_1 x_1) + (y_2 - w_0 - w_1 x_2) + \cdots + (y_n - w_0 - w_1 x_n)\right]$$

$$= -\frac{2}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i) = -\frac{2}{n}\left(\sum_{i=1}^{n} y_i - n w_0 - w_1 \sum_{i=1}^{n} x_i\right) = 2(w_0 + w_1 \bar{x} - \bar{y})$$

$$\frac{\partial E}{\partial w_1} = 0 = -\frac{2}{n}\left[x_1(y_1 - w_0 - w_1 x_1) + x_2(y_2 - w_0 - w_1 x_2) + \cdots + x_n(y_n - w_0 - w_1 x_n)\right]$$

$$= -\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - w_0 - w_1 x_i) = -\frac{2}{n}\left[\sum_{i=1}^{n}(x_i y_i) - w_0\sum_{i=1}^{n} x_i - w_1 \sum_{i=1}^{n} x_i^2\right] = 2\left(w_0 \bar{x} + w_1 \overline{x^2} - \overline{xy}\right)$$

$$\frac{\partial E}{\partial w_0} = 0 = 2\left(w_0 + w_1\bar{x} - \bar{y}\right)$$

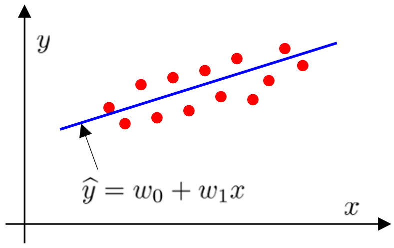
$$\frac{\partial E}{\partial w_1} = 0 = 2\left(w_0\bar{x} + w_1\overline{x^2} - \overline{xy}\right)$$

$$\boxed{\begin{aligned} w_0 + w_1\bar{x} &= \bar{y} \\ w_0\bar{x} + w_1\overline{x^2} &= \overline{xy} \end{aligned}} \qquad \boxed{\begin{aligned} w_0 &= \bar{y} - \bar{x}w_1 \\ w_1 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \end{aligned}}$$

$$w_0\bar{x} + w_1(\bar{x})^2 = \bar{x}\bar{y} \Rightarrow w_1\left[\overline{x^2} - (\bar{x})^2\right] = \overline{xy} - \bar{x}\bar{y}$$

Equations for finding the regression coefficients are called **Normal Equations**

Regression built the model: $\boxed{\widehat{y} = w_0 + w_1 x}$



Such regression analysis, involving linear relation between one dependent and one independent variables, is known as **Simple Linear Regression**

We can now estimate $\widehat{y}_\star = w_0 + w_1 x_\star$

which was the goal of our analysis

$$\boxed{\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i & \overline{x^2} &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i & \overline{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i) \end{aligned}}$$

averages

Alternative form of least square estimates w_0, w_1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$w_0 = \bar{y} - \bar{x}w_1$$

$$w_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \frac{1}{n}\sum_{i=1}^{n}(\bar{x}y_i)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})\,y_i\right]}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} = \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})\,y_i\right]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$S_{xy} = \sum_{i=1}^{n}\left[(x_i - \bar{x})\,y_i\right]$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}(\bar{x})^2 - 2\sum_{i=1}^{n}(x_i\bar{x}) = \sum_{i=1}^{n} x_i^2 + n(\bar{x})^2 - 2\bar{x}\sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i^2 + n(\bar{x})^2 - 2n(\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

We will now find an alternative form for S_{xy}

Again

$$\sum_{i=1}^{n}\left[(x_i-\bar{x})\,\bar{y}\right] = \bar{y}\sum_{i=1}^{n}(x_i-\bar{x}) = \bar{y}\sum_{i=1}^{n}x_i - \bar{y}\sum_{i=1}^{n}\bar{x}$$

$$= \bar{y}n\bar{x} - \bar{y}n\bar{x} = 0$$

Now

$$S_{xy} = \sum_{i=1}^{n}\left[(x_i-\bar{x})\,y_i\right] = \sum_{i=1}^{n}\left[(x_i-\bar{x})\,y_i\right] - \sum_{i=1}^{n}\left[(x_i-\bar{x})\,\bar{y}\right]$$

$$= \sum_{i=1}^{n}\left[(x_i-\bar{x})\,(y_i-\bar{y})\right]$$

Thus $\quad S_{xy} = \sum_{i=1}^{n}\left[(x_i-\bar{x})\,(y_i-\bar{y})\right]$

this form is sometimes useful

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

$$w_1 = \frac{S_{xy}}{S_{xx}} \qquad w_0 = \bar{y} - \bar{x}w_1$$

$$S_{xx} = \sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$S_{xy} = \sum_{i=1}^{n}\left[(x_i-\bar{x})\,y_i\right] = \sum_{i=1}^{n}\left[(x_i-\bar{x})\,(y_i-\bar{y})\right]$$

$$\boxed{\hat{y} = w_0 + w_1 x}$$

Finally, we verify the minimization condition

$$\frac{\partial^2 E}{\partial w_0^2}\frac{\partial^2 E}{\partial w_1^2} - \left(\frac{\partial^2 E}{\partial w_0 w_1}\right)^2 > 0 \quad \text{and} \quad \frac{\partial^2 E}{\partial w_0^2} > 0$$

$$\frac{\partial E}{\partial w_0} = -\frac{2}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)$$

$$\frac{\partial E}{\partial w_1} = -\frac{2}{n}\sum_{i=1}^{n}x_i\,(y_i - w_0 - w_1 x_i)$$

$$\frac{\partial^2 E}{\partial w_0^2} = -\frac{2}{n}\sum_{i=1}^{n}(-1) = 2 \qquad \frac{\partial^2 E}{\partial w_1^2} = -\frac{2}{n}\sum_{i=1}^{n}x_i\,(-x_i) = \frac{2}{n}\sum_{i=1}^{n}x_i^2 \qquad \frac{\partial^2 E}{\partial w_0 \partial w_1} = -\frac{2}{n}\sum_{i=1}^{n}(-x_i) = 2\bar{x}$$

$$\frac{\partial^2 E}{\partial w_0^2}\frac{\partial^2 E}{\partial w_1^2} - \left(\frac{\partial^2 E}{\partial w_0 \partial w_1}\right)^2 = \frac{4}{n}\sum_{i=1}^{n}x_i^2 - 4\,(\bar{x})^2 = \frac{4}{n}\left[\sum_{i=1}^{n}x_i^2 - n\,(\bar{x})^2\right] = \frac{4}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 > 0$$

(proved before)

The minimization condition is, therefore, satisfied

$E\,(w_0, w_1)$ has only one stationary point which is the minimum (convex function)

Such property of the function $E\,(w_0, w_1)$ justifies the choice of cost function