8. Probabilistic Analysis of Theoretical Linear Regression

Note: The following problems are presented as a derivation in the video that follows. We encourage you to attempt them before watching the video.

Derivation of Theoretical Linear Least Squares Regression I

2/2 points (graded)

Normally, we should be thinking of linear regression being performed on a data set $\{(x_i, y_i)\}_{i=1}^n$, which we think of as *a deterministic collection of points in the Euclidean space*. It is helpful to also consider an idealized scenario, where we assume that X and Y are random variables that follow some joint probability distribution and they have finite first and second moments. In this problem, we will derive the solution to the **theoretical linear regression** problem.

Assume Var $(X) \neq 0$. The **theoretical linear (least squares) regression** of Y on X prescribes that we find a pair of real numbers a and b that minimize $\mathbb{E}\left[(Y-a-bX)^2\right]$, over all possible choices of the pair (a, b).

To do so, we will use a classical calculus technique. Let $f(a, b) = \mathbb{E}\left[(Y-a-bX)^2\right]$, and now we solve for the critical points where the gradient is zero.

Hint: Here, assume you can switch expectation and differentiation with respect to a and b. That is, $\partial_a \mathbb{E}\left[(\cdots)\right] = \mathbb{E}\left[\partial_a(\cdots)\right]$.

Use **X** and **Y** for random variables X and Y.

The partial derivatives are:

$\partial_a f = \mathbb{E}\Big[$ -2*(Y-a-b*X) ✔ **Answer:** -2*Y+2*a+2*b*X $\Big]$

$\partial_b f = \mathbb{E}\Big[$ -2*X*(Y-a-b*X) ✔ **Answer:** -2*X*Y + 2*a*X + 2*b*X^2 $\Big]$

STANDARD NOTATION

Solution:

As suggested, it's easier to take the derivative inside the expectation. Such a step is valid, since the expectation is an integral with respect to x and y, while the derivatives are taken with respect to a and b. In fact, keep in mind that we are differentiating with respect to a and b, so X and Y should be treated as constants. Using the chain rule, we obtain

$$\partial_a f = \mathbb{E}\left[-2\left(Y-a-bX\right)\right] = \mathbb{E}\left[-2Y+2a+2bX\right]$$

$$\partial_b f = \mathbb{E}\left[-2X\left(Y-a-bX\right)\right] = \mathbb{E}\left[-2XY+2aX+2bX^2\right].$$

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Derivation of Theoretical Linear Least Squares Regression II

1/1 point (graded)

Setting these equal to zero and isolating terms with a and b to one side, we obtain a system of linear equations

$$\boxed{\begin{aligned}\mathbb{E}[Y] &= a + \mathbb{E}[X]\, b \\ \mathbb{E}[XY] &= \mathbb{E}[X]\, a + \mathbb{E}[X^2]\, b\end{aligned}}$$

Multiplying the first equation by $\mathbb{E}[X]$ and subtracting from the second equation gives

$$\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) b = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \quad \Longrightarrow \quad \boxed{b = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}}.$$

Plugging this value back into the first equation to solve for a gives

$$\boxed{a = \mathbb{E}[Y] - \frac{\text{Cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]}.$$

We now compute the Hessian

$$H = \begin{pmatrix} f_{aa} & f_{ab} \\ f_{ba} & f_{bb} \end{pmatrix}$$

to make sure that this pair (a, b) critical point is a local minimum. The determinant of H at this value (a, b) is

- $-\text{Var}(X)$
- $4\text{Var}(X)$ ✔
- $\mathbb{E}[X]$
- $\text{Cov}(X,Y)$

Solution:

The second derivatives can be evaluated using the answers from the first problem, which were:

$$\partial_a f = \mathbb{E}[-2Y + 2a + 2bX], \qquad \partial_b f = \mathbb{E}[-2XY + 2aX + 2bX^2].$$

We demonstrate how to compute $\partial_{aa} f$ as follows:

$$\begin{aligned}\partial_{aa} f &= \partial_a(\partial_a f) \\ &= \mathbb{E}[\partial_a(-2Y + 2a + 2bX)] \\ &= \mathbb{E}[2] = 2.\end{aligned}$$

Similarly, the Hessian evaluates to the matrix

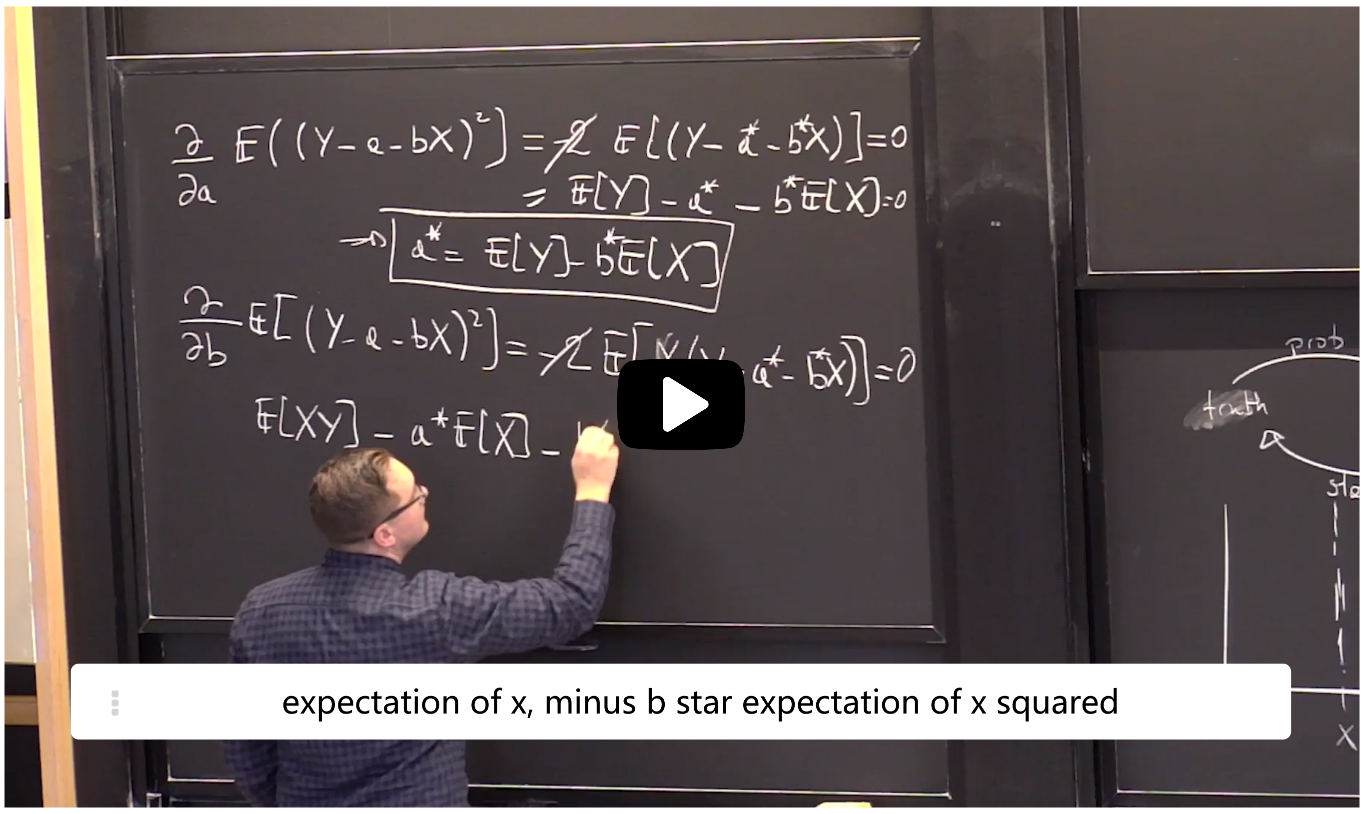
$$H = \begin{pmatrix} 2 & 2\mathbb{E}[X] \\ 2\mathbb{E}[X] & 2\mathbb{E}[X^2] \end{pmatrix}$$

which has determinant $4\mathbb{E}[X^2] - 4\mathbb{E}[X]^2 = 4\text{Var}(X)$, independent of the value of a and b, and is always positive. This justifies the positive-definiteness of the Hessian everywhere, which means that f is **strictly convex**. Therefore, (a, b) is the global minimizer of f.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Optimal Theoretical Regression Line



5:43 / 9:29 1.0x

and so it says the expectation of xy minus a star

expectation of x, minus b star expectation of x squared

is equal to 0.

So now I have to solve for those two guys--

so I'm just going to plug in my a star

equal to this expectation, OK?

So that tells me the expectation of xy is equal to expectation--

sorry, minus expectation of x expectation of y.

So I'm just replacing here by expectation of y,

and then I have plus b star expectation of x, and there's

two of them, and one that's coming from this guy,

and one that's coming from that guy--

minus b star expectation of x squared equals 0.

Now you can start seeing that there's things

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Theoretical Linear Regression Visualized I

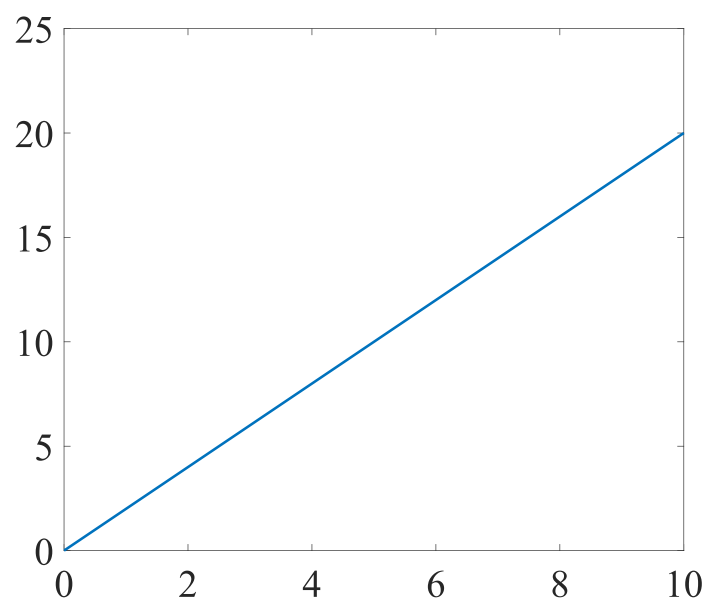
1/1 point (graded)

Consider again the setting of theoretical linear regression, as in the previous problems on this page. Let X, Y be random variables such that $\mathsf{Var}\,(X) \neq 0$. Assume $\mathbb{E}\,[X]$ and $\mathbb{E}\,[Y]$ are both zero.

Let a, b be solutions that minimze the squared error

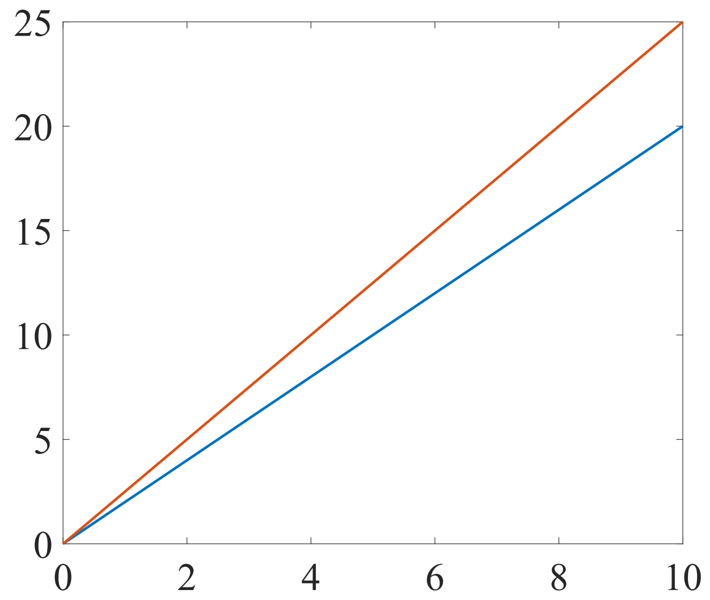
$$a = \mathbb{E}\,[Y] - \frac{\mathsf{Cov}\,(X,Y)}{\mathsf{Var}\,(X)}\mathbb{E}\,[X]\,, \qquad b = \frac{\mathsf{Cov}\,(X,Y)}{\mathsf{Var}\,(X)}$$

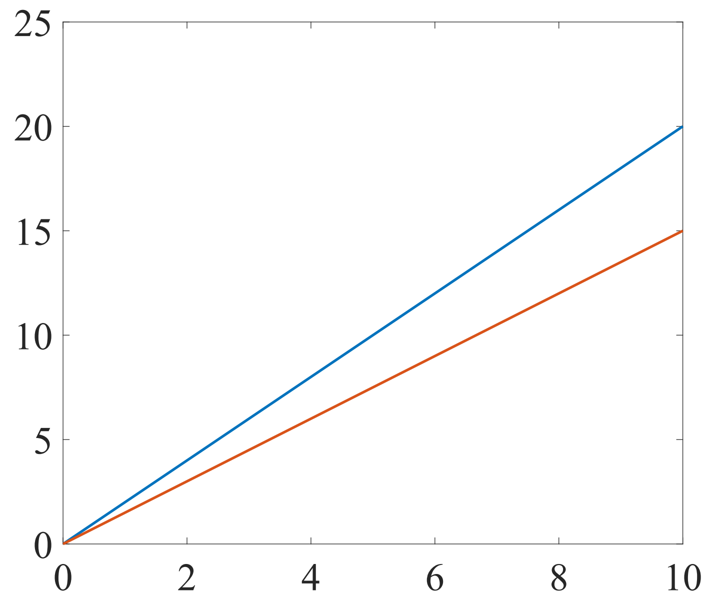
which gives the best-fitting line $\mathbb{E}\,[Y|X=x] \approx a + bx$. Assume that the line $y = a + bx$ looks like:



In particular, $a = 0$ due to our simplifying assumptions.

If Y' is a different random variable such that $\mathbb{E}[Y'] = 0$, Cov $(X, Y') >$ Cov (X, Y), which of the following choices best illustrates, via a new line drawn in red, the theoretical linear regression of the pair X, Y'?





Solution:

Increasing the covariance increases b and hence the slope increases. Qualitatively, the reason why the slope *ought to increase* if the covariance increases is revealed in the definition of covariance:

$$\text{Cov}\,(X, Y) = \mathbb{E}[\sum_{y \in \Omega_Y} (x - \mathbb{E}[X])\,(y - \mathbb{E}[Y])]$$

If X is held fixed, then the covariance increases if, for each x:

1. $y - \mathbb{E}[Y]$ tends to be more positive whenever $x > \mathbb{E}[X]$, and
2. $y - \mathbb{E}[Y]$ tends to be more negative whenever $x < \mathbb{E}[X]$.

Which, in our scenario, means that for a typical sample (x, y), the y-coordinate tends to be more positive on average whenever $x > 0$, and more negative whenever $x < 0$.

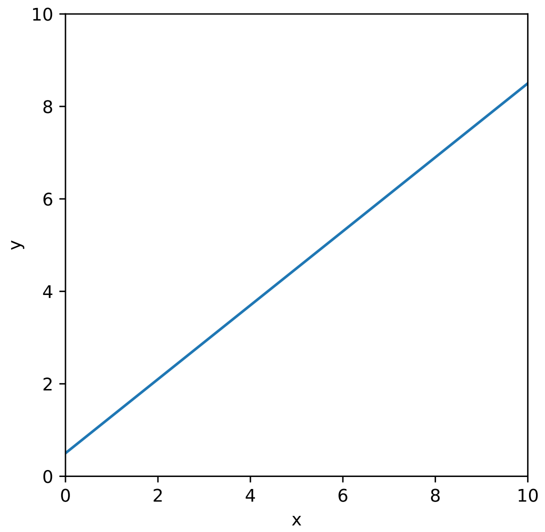
Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Theoretical Linear Regression Visualized II

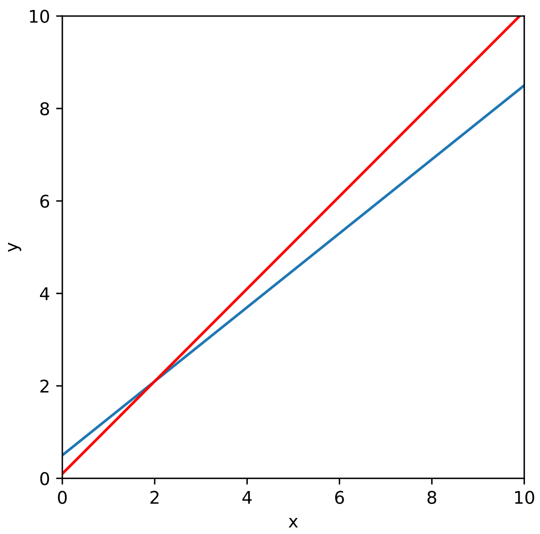
1/1 point (graded)
Now consider the same setting as in the previous problem, except we drop the assumption $\mathbb{E}[Y] = 0$, and we now assume $\mathbb{E}[X] > 0$.

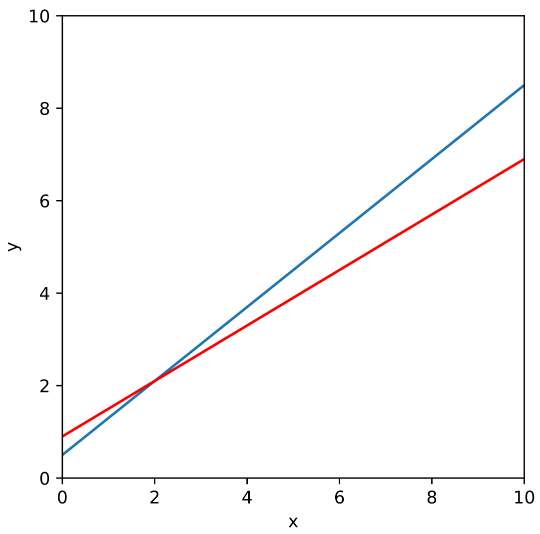
Again, let a, b be solutions that minimze the squared error, so that the line $y = a + bx$ looks like:

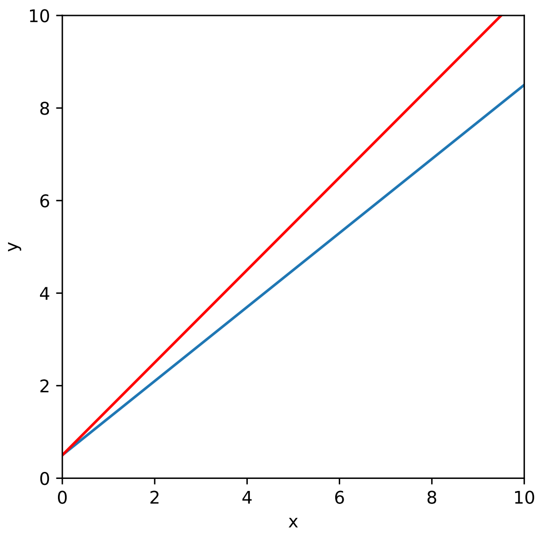


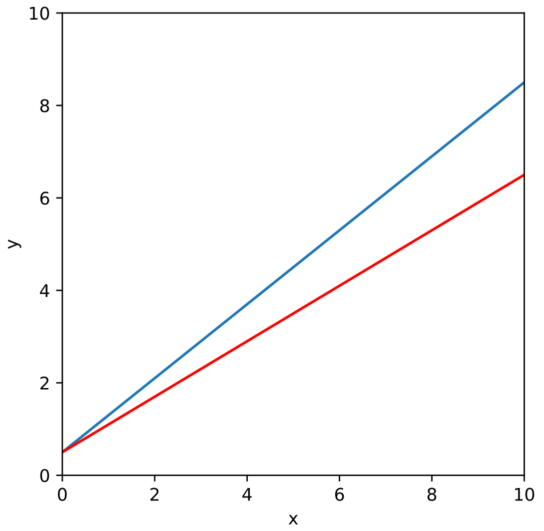
新的更小

If Y' is a different random variable such that $\text{Cov}\left(X, Y'\right) > \text{Cov}\left(X, Y\right)$ and $\mathbb{E}\left[Y\right] \geq \mathbb{E}\left[Y'\right]$, which of the following choices best illustrates, via a new line drawn in red, the theoretical linear regression of the pair X, Y'?









Solution:

The reasoning is almost the same as the previous problem's, with an extra step. The slope increases while the intercept decreases.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Assumptions of Theoretical Linear Regression

1/1 point (graded)

Let us think about what goes wrong when we drop the assumption that Var $(X) \neq 0$ *in theoretical linear regression.*

Let X and Y be two real random variables with two moments, and Var $(X) = 0$. (**Note:** the variance of X is zero whenever $\mathbf{P}(X = \mathbb{E}[X]) = 1$.) We make no further assumptions on Y.

Which one of the following statements is **false**?

- There is an infinite family of solutions (a, b) that minimize the squared mean error, $\mathbb{E}\left[(Y - a - bX)^2\right]$.
- There is no line $y = a + bx$ that predicts $\boxed{Y \text{ given } X}$ with probability 1, regardless of their distribution. ✔
- With probability equal to 1, the random pair (X, Y) lies on the vertical line $x = \mathbb{E}[X]$.

Solution:

First, a technical remark: one might be tempted to say that the "best fitting line" is the vertical line $x = \mathbb{E}[X]$. However, this is not in the family of lines $y = a + bx$ parametrized by (a, b), which is what the first two choices are asking about.
The only false statement here is "Y **can never be predicted from** X.". We analyze the choices one by one. For convenience, let $x_0 = \mathbb{E}[X]$.

- "**There is an infinite family of solutions** (a, b) **that minimize the squared mean error,** $\mathbb{E}\left[(Y - a - bX)^2\right]$." There are, indeed, an infinite family of lines from the family $y = a + bx$ that minimize the mean squared error. Since Var $(X) = 0$, we have

$$\mathbb{E}_{X,Y}\left[(Y - a - bX)^2\right] = \mathbb{E}_Y\left[(Y - a - bx_0)^2\right]$$

Introduce the variable $c = a + bx_0$, which represents the predicted y-coordinate at $x = x_0$. This simplifies the above expectation to $\mathbb{E}_Y\left[(Y - c)^2\right]$, which is minimized when $c = \mathbb{E}[Y]$. This tells us the following important fact: *any line $y = a + bx$ which crosses the point $(x_0, \mathbb{E}[Y])$ minimizes the mean error.* 任何斜率的都可以

- "**There is no line $y = a + bx$ that predicts Y given X with probability 1, regardless of their distribution.**" This is false; consider the case where Var (Y) is also zero. Then we can make a prediction for Y, which is simply $\mathbb{E}[Y]$. By the same reasoning about Var $(X) = 0$ whenever all of the likelihood is concentrated on a single point, this prediction is correct with probability 1.

- "**With probability equal to 1, the random pair (X, Y) lies on the vertical line $x = x_0$.**" This is true, because it is simply a re-statement of the remark made in the problem statement: $\mathbf{P}(X = \mathbb{E}[X]) = 1$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 6 Linear Regression:Lectures 19: Linear Regression 1 / 8. Probabilistic Analysis of Theoretical Linear Regression