11. Multivariate Regression: Definitions, Modeling, and Matrix LSE

Bookmark this page

Exercises due Aug 24, 2021 19:59 EDT Completed

Multivariate Regression: Setup and Definitions

Start of transcript. Skip to the end.

0:00 / 0:00 1.50x

So multivariate regression is the case where

I have more than just one x.

So we mentioned that, for example, you

might want to predict your total conversion

rates by a combination of maybe the number of clicks

and the amount spent on your campaign

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

LSE in Matrix Form: Setup

Start of transcript. Skip to the end.

0:00 / 0:00 1.50x

OK.

So it turns out that there's a much more concise way

of writing the sum of squares.

So sum of squares, when you start talking about vectors,

should actually immediately beg to use squared norm, right?

Just like we had the chi squared,

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

The **multivariate linear model** can be described via the equation $Y = \mathbf{X}^T\boldsymbol{\beta} + \varepsilon$, where:

- $\mathbf{X} \in \mathbb{R}^p$ is the vector of **covariates** , also called **independent/explanatory** variables,
- $Y \in \mathbb{R}$ is the **dependent** variable,
- $\varepsilon \in \mathbb{R}$ is the noise, and
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is the model parameter

(**Note:** We may have also written $\boldsymbol{\beta}^T \mathbf{X}$ instead of $\mathbf{X}^T \boldsymbol{\beta}$. These are transposes of each other, but they are equal since they are both scalars. Recall that the transpose of a scalar is itself.)

If we have n observations $\{(\mathbf{X}_i, Y_i)\}$, then this determines n linear relationships, each of the form $Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i$. We can stack these into a matrix equation:

$$\begin{aligned} Y_1 &= \mathbf{X}_1^T \boldsymbol{\beta} + \varepsilon_1 \\ Y_2 &= \mathbf{X}_2^T \boldsymbol{\beta} + \varepsilon_2 \\ &\vdots \\ Y_n &= \mathbf{X}_n^T \boldsymbol{\beta} + \varepsilon_n \end{aligned} \quad \leftrightarrow \quad \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{10.1}$$

In this course, we typically condense the equation on the right into the form $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

“Model" versus “Regression":

The assumption that the random variable pair $(\mathbf{X}, Y)$ obeys the relationship $Y = \mathbf{X}^T \boldsymbol{\beta} + \varepsilon$ is an assumption on the *model*. Equivalently, we can assume that the regression function is linear: $\mu(x) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}] = \mathbf{x}^T \boldsymbol{\beta}$, with the understanding that $\mathbb{E}[\varepsilon] = 0$.

This allows us to perform **linear regression** , which consists of coming up with an estimator $\hat{\boldsymbol{\beta}}$ in an attempt to find the best-fitting guess $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$.

Note that we can always **perform** linear regression, even if the model is misspecified. There are many ways that things can go wrong! For example, the estimator may not be unique, or the estimator $\boldsymbol{\beta}$ may have huge variance. **This unit will help us understand when and why these issues occur.**

How does this relate to the single-variable setting?

Recall that in the previous section ($p = 1$), the model was $Y = a + bX + \varepsilon$ for scalar values of a, b, X, Y, ε. To write this down using the notation in the multivariate setting, take

$$\beta = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} 1 \\ X \end{pmatrix}.$$

To extrapolate from the single-variable case, consider the p-dimensional linear model with intercept β_0 which looks like

$$Y = \beta_0 + \beta_1 X^{(1)} + \beta_2 X^{(2)} + \cdots + \beta_p X^{(p)} + \varepsilon.$$

The natural analogy is to take $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T \in \mathbb{R}^{p+1}$ and $\mathbf{X} = (1, X^{(1)}, \ldots, X^{(p)}) \in \mathbb{R}^{p+1}$. Therefore, whenever we have an intercept in the model, we extend the dimension by 1 and take the first coordinate of $\mathbf{X}$ to always be 1.

(On the other hand, if we did not have an intercept in our model, then we would not need β_0. In this case, for a typical p-dimensional model, we usually write $\mathbf{X} = (X^{(1)}, \ldots, X^{(p)})$, a p-dimensional vector.)

This technical distinction won't affect theoretical analyses. **Unless otherwise specified, we will always take $\mathbf{X}$ and $\boldsymbol{\beta}$ to be generic vectors in $\mathbb{R}^p$.**

Linear Regression as a Statistical Model I

2/2 points (graded)
Consider the linear regression model introduced in the slides and lecture, restated below:

Linear regression model : $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$ are i.i.d from the linear regression model $Y_i = \boldsymbol{\beta}^\top \mathbf{X}_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$ for an unknown $\boldsymbol{\beta} \in \mathbb{R}^d$ and $\mathbf{X}_i \sim \mathcal{N}_d(0, I_d)$ independent of ε_i.

Suppose that $\boldsymbol{\beta} = \mathbf{1} \in \mathbb{R}^d$, which denotes the d-dimensional vector with all entries equal to 1.

What is the mean of Y_1?

$\mathbb{E}[Y_1] =$

What is the variance of Y_1? (Express your answer in terms of d.)

Var $(Y_1) =$

STANDARD NOTATION

Submit You have used 1 of 2 attempts

Linear Regression as a Statistical Model II

2/2 points (graded)
Recall the linear regression model as introduced above in the previous question. This model is parametric, although it is not written in the standard notation previously introduced for parametric statistical models. In this problem, you will explicitly write the linear regression model as a parametric statistical model.

We will represent the linear regression model as an ordered pair $(E, \{P_\beta\}_{\beta\in\Theta})$. Here E denotes the sample space associated to the distribution P_β, where P_β is defined as follows for $\beta \in \mathbb{R}^d$:

The random ordered pair $(\mathbf{X}, Y) \subset \mathbb{R}^d \times \mathbb{R}$ *is distributed as* P_β *if:*

- $\mathbf{X} \sim \mathcal{N}(0, I_d)$,
- $Y \sim \beta^T X + \varepsilon$, *where* $\varepsilon \sim \mathcal{N}(0,1)$ *and* ε *is independent of* $\mathbf{X}$.

The set Θ in the ordered pair $(E, \{P_\beta\}_{\beta\in\Theta})$ denotes the parameter space for this model.

The sample space for the linear regression model can be written $E = \mathbb{R}^k$ for some integer k. What is k? (Express your answer in terms of d.)

Hint: You should use the fact that $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$ for all integers $m, n \geq 0$.

$k =$

The parameter space for the model can be written as $\Theta = \mathbb{R}^j$ for some integer j. What is j? (Express your answer in terms of d.)

$j =$

STANDARD NOTATION

Submit You have used 2 of 2 attempts

Discussion

Hide Discussion

Topic: Unit 6 Linear Regression:Lecture 19: Linear Regression 1 / 11. Multivariate Regression: Definitions, Modeling, and Matrix LSE

Add a Post

Show all posts by recent activity

Here we replaced conditional expectation by an average using the statistical hammer, but where is 1/n ? 3

? How does the answer to the last question change if we had n observations? 2

Previous Next