

<u>Help</u>

HuitianDiao 🗸

<u>Course</u> <u>Progress</u> <u>Dates</u> <u>Discussion</u> <u>Resources</u>

★ Course / Unit 6 Linear Regression / Lecture 19: Linear Regression 1





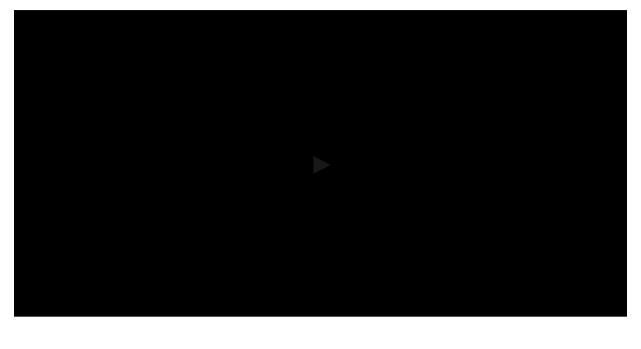
11. Multivariate Regression: Definitions, Modeling, and Matrix LSE

☐ Bookmark this page

Exercises due Aug 24, 2021 19:59 EDT Completed

Multivariate Regression: Setup and Definitions

Start of transcript. Skip to the end.



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

66

LSE in Matrix Form: Setup

Start of transcript. Skip to the end.



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

<u>Download SubRip (.srt) file</u> <u>Download Text (.txt) file</u>

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,
- $oldsymbol{arepsilon} \in \mathbb{R}$ is the noise, and
- $\mathbf{A} \in \mathbb{R}^p$ is the model parameter

b c me include 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:

$$Y_{1} = \mathbf{X}_{1}^{T}\boldsymbol{\beta} + \varepsilon_{1} Y_{2} = \mathbf{X}_{2}^{T}\boldsymbol{\beta} + \varepsilon_{2} \vdots Y_{n} = \mathbf{X}_{n}^{T}\boldsymbol{\beta} + \varepsilon_{n}$$
 \leftrightarrow
$$\begin{pmatrix} Y_{1} \\ Y_{2} \\ \vdots \\ Y_{n} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1}^{T} \\ \mathbf{X}_{2}^{T} \\ \vdots \\ \mathbf{X}_{n}^{T} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_{1} \\ \varepsilon_{2} \\ \vdots \\ \varepsilon_{n} \end{pmatrix}$$
 (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: $\mu\left(x\right)=\mathbb{E}\left[Y|\mathbf{X}=\mathbf{x}\right]=\mathbf{x}^T\boldsymbol{\beta}$, with the understanding that $\mathbb{E}\left[\varepsilon\right]=0$.

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

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 β 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

$$eta = \left(egin{array}{c} a \ b \end{array}
ight), \qquad \mathbf{x} = \left(egin{array}{c} 1 \ X \end{array}
ight).$$

To extrapolate from the single-variable case, consider the p-dimensional linear model with intercept eta_0 which looks like

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

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 $\mathbf{1}$ and take the first coordinate of \mathbf{X} to always be $\mathbf{1}$.

(On the other hand, if we did not have an intercept in our model, then we would not need eta_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 X and β 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:

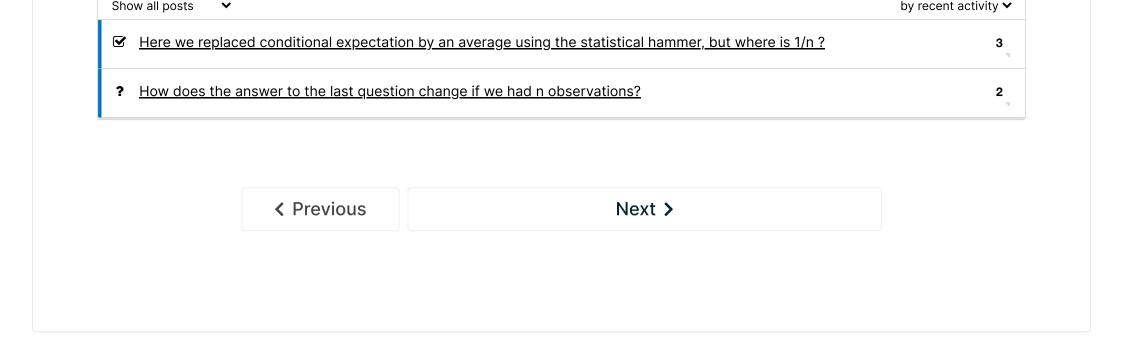
Linear regression model : $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$ are i.i.d from the linear regression model $Y_i = \boldsymbol{\beta}^\top \mathbf{X}_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0,1\right)$ for an unknown $\boldsymbol{\beta}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

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

what is the mean of Y_1 ?
$\mathbb{E}\left[Y_1 ight] =$
What is the variance of Y_1 ? (Express your answer in terms of $oldsymbol{d}$.)
$Var\left(Y_{1}\right) =$
CTANDARD NOTATION
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 $(E,\{P_eta\}_{eta\in\Theta})$. Here E denotes the sample space associated to the distribution P_eta , where P_eta is defined as follows for $m{eta}\in\mathbb{R}^d$:
The random ordered pair $(\mathbf{X},Y)\subset \mathbb{R}^d imes \mathbb{R}$ is distributed as $P_{m{eta}}$ if:
$oldsymbol{\cdot} \; \mathbf{X} \sim \mathcal{N}\left(0, I_d ight),$
$m{\cdot}\; Y\sim m{eta}^TX+arepsilon$, where $arepsilon\sim \mathcal{N}\left(0,1 ight)$ and $arepsilon$ is independent of $m{X}$.
The set Θ in the ordered pair $(E,\{P_{m{eta}}\}_{m{eta}\in\Theta})$ 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 $\mathbb{R}^{m+n}=\mathbb{R}^m imes\mathbb{R}^n$ 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 $m{j}$. What is $m{j}$? (Express your answer in terms of $m{d}$.)
j =
STANDARD NOTATION
Submit You have used 2 of 2 attempts
Discussion

Discussion

Hide Discussion



© All Rights Reserved



edX

About

Affiliates

edX for Business

Open edX

<u>Careers</u>

<u>News</u>

Legal

Terms of Service & Honor Code

Privacy Policy

Accessibility Policy

Trademark Policy

<u>Sitemap</u>

Connect

Blog

Contact Us

Help Center

Media Kit

Donate















© 2021 edX Inc. All rights reserved.

深圳市恒宇博科技有限公司 <u>粤ICP备17044299号-2</u>