Discussion 8: Linear Regression and Modeling

Data 100, Spring 2020

Suraj Rampure

Friday, March 13th, 2020

Agenda

- Multiple Regression
- Solution to OLS
- Features

As per usual, everything will be posted at

surajrampure.com/teaching/ds100.html

Multiple Regression

Simple Linear Regression

$y = a + bx$

Our model is

intercept

$\theta^T x = \hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_p x_p$

p features

- $x_1, x_2, \ldots, x_p$ are called "features", "covariates", "explanatory variables", columns"
- $\theta_0, \theta_1, \ldots, \theta_p$ are called "weights"
- y is called the "response variable"

$\hat{y}$ is the predicted version of y

Our goal is to model the relationship between explanatory variables and our response variable. More concretely, our goal is to find values of $\theta_0, \theta_1, \ldots, \theta_p$ that minimize our average L_2 loss:

L_2 : (actual − pred)^2

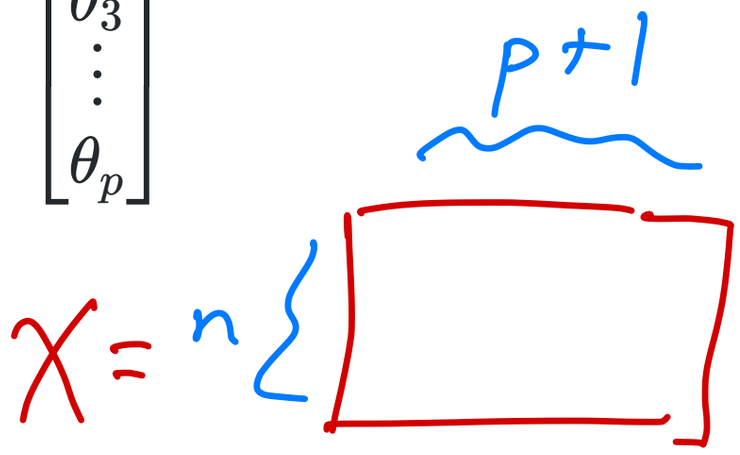
$$L(\vec{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

We denote the optimal values by $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_p$, or more generally, by $\hat{\theta}$.

Matrix Formulation

Our model is $\hat{y} = X\theta$:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \cdots & x_{3p} \\ & & & & \vdots & \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_p \end{bmatrix}$$



- n data points (observations). Each corresponds to one row.
- p features, plus a bias column. Each column is a "feature".
- Almost always, $n > p$.

Under this formulation...

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}||y - X\theta||_2^2$$ → L_2

Question: How do we find the optimal $\hat{\theta}$?

From the previous slide, our objective function is the average L_2 loss over our entire dataset (i.e. the mean squared error, MSE).

$$L(\theta) = \frac{1}{n}||y - X\theta||_2^2$$

There are two ways to solve for $\hat{\theta}$:

1. Through calculus: take the gradient, set it equal to 0.
    - We are not doing this in our semester, but just know that it can be done.
2. Through geometry.
    - This was done in lecture, and we will recap the argument in the worksheet.

Either way, we end up at the same solution:

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

Using our Model

dot prod: $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$, $a \cdot b = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$

Given our initial data X and y, we now know how to solve for the vector $\hat{\theta}$ that minimizes our model's loss. Now, given this vector, how do we use it to make predictions for **unseen data**?

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ \text{height} \\ \text{weight} \end{bmatrix}$$

$$\hat{\theta} \cdot x = \hat{\theta}_0 + \hat{\theta}_1 \cdot \text{height} + \hat{\theta}_2 \cdot \text{weight}$$

$\hat{\theta} \cdot x$, $\hat{\theta}^T x$, $x^T \hat{\theta}$

→ all equivalent!

Features

Note: In order for us to perform linear regression, our model only needs to be **linear in terms of the weights**.

For example, let's suppose we want to predict `mpg` given `horsepower` (x_1), `model year` (x_2) and `acceleration` (x_3). The following are both linear:

$$\rightarrow \quad y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$\rightarrow \quad y = \theta_0 + \theta_1 x_1 + \theta_2 \underbrace{x_1^2 \sin x_3}_{x_4} + \theta_3 \underbrace{e^{x_1 x_2 x_3}}_{x_5}$$

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_4 + \theta_3 x_5$

On the other hand, the following model is **not linear**: why?

$$y = \theta_0 + \theta_1 x_1^{\theta_2} + \theta_2^2 x_1 x_2 x_3$$

could not model with $\hat{y} = X\theta$

$$X = \begin{bmatrix} 1 & x_{11} & (x_1^2 \sin x_3)_1 & (e^{x_1 x_2 x_3})_1 \\ 1 & x_{21} & (x_1^2 \sin x_3)_2 & (e^{x_1 x_2 x_3})_2 \\ 1 & x_{31} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & (x_1^2 \sin x_3)_n & \end{bmatrix}$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 \sin x_3 + \theta_3 e^{x_1 x_2 x_3}$$