Least Squares Linear Regression

Raguvir Kunani

Discussion 9

October 23, 2019

Simple Linear Regression Quick Review

Simple linear regression involves finding a "line of best fit" that explains the relationship between **2** variables x and y. In fancy math terms:

$$\min_{a,b} \sum_{i=1}^{n} \underbrace{(y_i - (a + bx_i))}_{\text{residual}}^{2}$$

L2 loss

sum

All this really means is we are trying to find the values of a and b so that a line with the equation $\hat{y} = a + bx$ best fits the data.

Residuals

The **residual** is a measure of how "good" our prediction was. In simple linear regression, the residual for the i^{th} point is:

$$e_i = \underbrace{y_i}_{\text{true value}} - (\underbrace{a + bx_i}_{\text{prediction}})$$

So really what linear regression is doing is finding the a and b that lead to the minimum sum of residuals. This view of linear regression is very helpful to understanding linear regression in higher dimensions.

Linear Algebra Review: Vectors

A **vector** is an ordered list of numbers. An n-dimensional vector contains n numbers. A fancy math notation for saying a vector x is an n-dimensional vector is $x \in \mathbb{R}^n$.

Vectors are assumed to be column vectors unless otherwise stated.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Linear Algebra Review: Matrices

A **matrix** is a rectangular array of numbers. It often helps to think of a matrix as a collection of vectors. Every matrix has 2 dimensions: the number of rows m and the number of columns n. If a matrix A has m rows and n columns, we say that $A \in \mathbb{R}^{m \times n}$.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

In general, $m \neq n$. But when $m = n$, we say A is a square matrix.

Linear Algebra Review: Vector Tranposes

The transpose of a column vector is a row vector and vice versa.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$x^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Linear Algebra Review: Matrix Tranposes

The tranpose of a matrix flips the rows and columns.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

$$A^T = \begin{bmatrix} A_{11} & \cdots & A_{m1} \\ \vdots & \ddots & \vdots \\ A_{1n} & \cdots & A_{mn} \end{bmatrix}$$

Linear Algebra Review: Vector and Matrix Products

Let $x, y \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$.

Vector Inner Product: $x^T y$ or $y^T x$

Dot Product $x \cdot y$ $y \cdot x$

Matrix-Vector Product: Ax

Closer Look at Matrix-Vector Product

A matrix vector product Ax can be thought of as a linear combination of the columns of A, which is often a useful interpretation.

$$\underset{A}{\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}} \underset{x}{\begin{bmatrix} 1 \\ 2 \end{bmatrix}} = \underset{y}{\begin{bmatrix} 11 \\ 17 \end{bmatrix}} = 1 \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 2 \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

linear combo of
columns of A!

Orthogonality of Vectors / Spaces

Two vectors x and y are **orthogonal** if $x^T y = 0$. This definition can be extended to spaces: x is orthogonal to $\text{span}(A)$ if $x^T A y = 0 \; \forall y$.

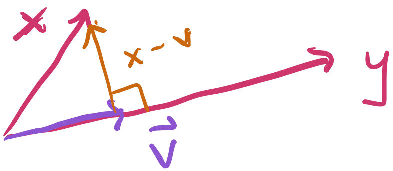
Why do we care about this for linear regression?

↳ we'll see this later, but it has to do with projections

Projections

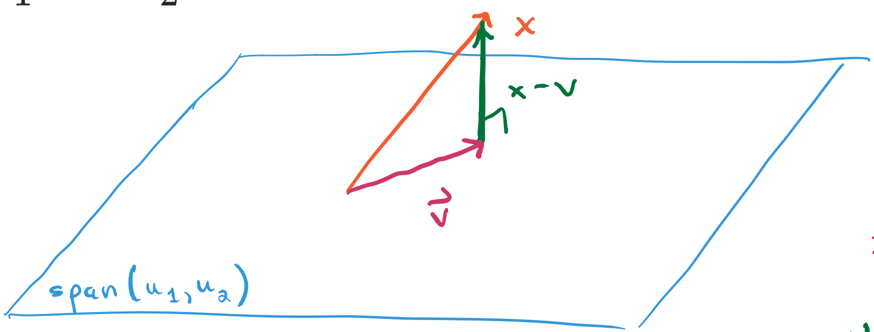
x - v orthogonal to y!

The projection of a vector x onto a vector y is the vector *in the same direction as* y that is closest in distance to x.



$\vec{v}$ is the projection of x onto y

The projection of a vector x onto $\text{span}(u_1, u_2)$ is the vector in the span of u_1 and u_2 that is closest in distance to x.



$\vec{v}$ is projection of x onto span(u_1, u_2)

$x - v$ orthogonal to span(u_1, u_2)

Euclidean Norm as Distance

In this class, we will always use the Euclidean norm (aka 2-norm) as the measure of the distance between 2 vectors. The (squared) distance between x and y is written as:

squared

$$||x - y||_2^2 = \sum_{i=1}^{n} (x_i - y_i)^2$$

2 - norm

Notice that this looks a lot like a residual! This is not a coincidence.

Least Squares Linear Regression in N dimensions

In this class, we formulate least squares linear regression as:

>2D

$$\min_{\beta} ||y - X\beta||_2^2$$

true values

data

coefficients / weights

Compare this to the simple linear regression formulation:

2D

data points

$$\min_{a,b} \sum_{i=1}^{n} \underbrace{(y_i - (a + bx_i))}^2$$

$$\left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2^2$$

Least Squares as a Projection

How in the world are we supposed to solve $\min_\beta ||y - X\beta||_2^2$?

Projections! We know that

1. $||y - X\beta||_2^2$ is the distance between y and $X\beta$
2. Projections are about finding minimum distance

So really what we are doing is finding the projection of y onto $\mathrm{span}(\mathrm{cols}(X))$!

Worksheet!

Feedback Form

This *anonymous* form is for me to learn what I can do to ensure you all get the most of discussion and lab. This form will be open all semester, and I'll be checking it regularly. Be as ruthless as you want, I promise my feelings won't get hurt.

Feedback Form: tinyurl.com/raguvirTAfeedback