medford-group / data_analytics_ChE

Code Issues Pull requests Actions Projects Security Insight

Join GitHub today

GitHub is home to over 50 million developers working together to host and review code, manage projects, and build software together.

Sign up

Dismiss

master

data_analytics_ChE / 1-numerical_methods / Topic3-Linear_Regression.ipynb

Sihoon Choi exercises updated

History

0 contributors

Download

2.19 MB

Table of Contents

Linear Regression

Simple linear regression

Linear regression is a great starting point for understanding how linear algebra and optimization are used together for data analytics. We will start with simple linear regression, which you should be familiar with.

The form of a simple linear regression model is given as:

$y = mx + b + \epsilon$

where the y is the independent value, x is the dependent value, m is the slope of the line, b is the intercept, and ϵ is the "error" between the model and the actual data. This can also be written with indices on the data:

$y_i = mx_i + b + \epsilon_i$

where i refers to the index of the data point (e.g. the first, second, third, ... data point). We can also think of these quantities as vectors:

$\vec{y} = m\vec{x} + b + \vec{\epsilon}$

To make things consistent with prior lectures, we can re-write this as:

$y_i = w_0 x_i^0 + w_1 x_i^1 + \epsilon_i$

where $w_0 = b$ and $w_1 = m$. Now we can re-write this as a matrix-vector product:

$y_i = \sum_{j=0}^{1} w_j x_i^j + \epsilon_i$

If you recall the Vandermonde matrix, this can be written as: