Linear Regression

Mohammad Emtiyaz Khan
EPFL

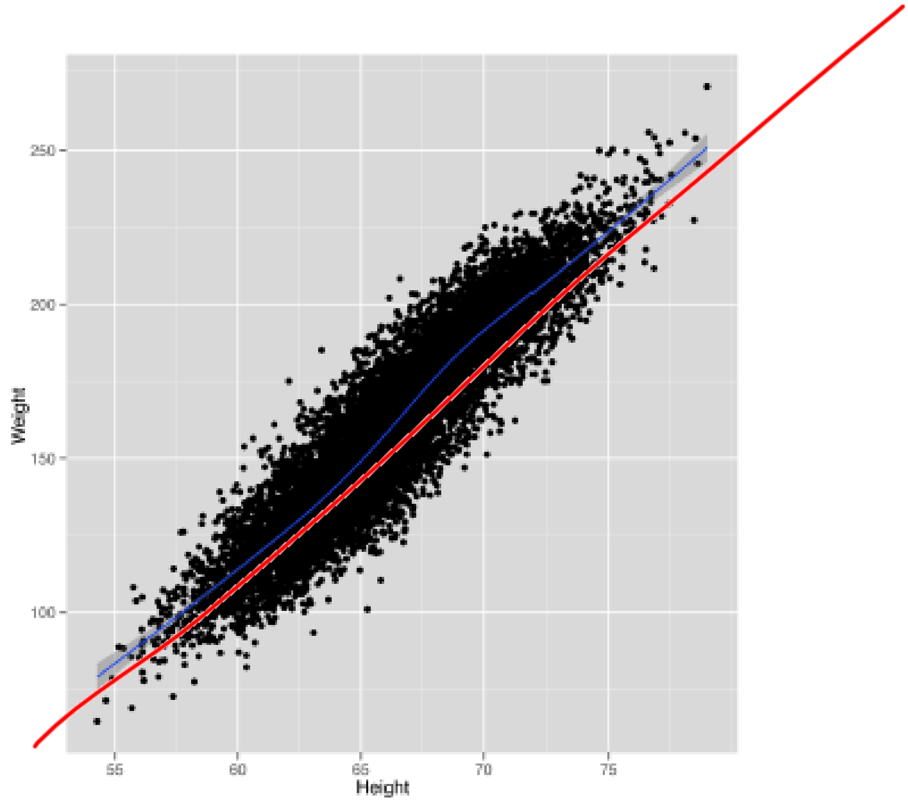
Sep 17, 2015

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

1 Model: Linear Regression

What is it?

Linear regression is a model that assumes a linear relationship between inputs and the output.



Why learn about *linear* regression?

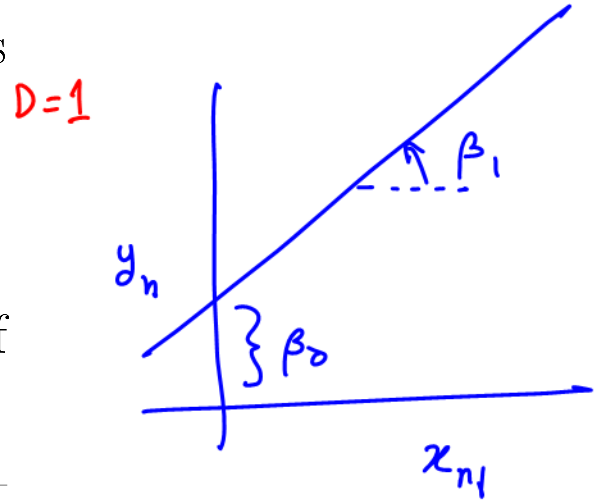
Plenty of reasons: simple, easy to understand, most widely used, easily generalized to non-linear models. Most importantly, you can learn almost all fundamental concepts of ML with regression alone.

Simple linear regression

With only one input dimension, it is simple linear regression.

$$y_n \approx f(\mathbf{x}_n) := \beta_0 + \beta_1 x_{n1}$$

Here, β_0 and β_1 are parameters of the model.



Multiple linear regression

With multiple input dimension, it is multiple linear regression.

$$\begin{aligned} y_n &\approx f(\mathbf{x}_n) \\ &:= \beta_0 + \beta_1 x_{n1} + \ldots + \beta_D x_{nD} \\ &= \tilde{\mathbf{x}}_n^T \boldsymbol{\beta} \end{aligned} \quad (1)$$

Define,

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_D \end{bmatrix} \qquad \underline{x}_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nD} \end{bmatrix} \qquad \tilde{\underline{x}}_n = \begin{bmatrix} 1 \\ \underline{x}_n \end{bmatrix}$$

Learning/estimation/fitting

Given data, we would like to find $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_D]$. This is called learning or estimating the parameters or fitting the model.

$$\mathcal{D} = \{y_n, \underline{x}_n\}_{n=1}^N$$

Additional Notes

Matrix multiplication

To go any further, one must revise matrix multiplication. Remember that multiplication of $M \times N$ matrix with a $N \times D$ matrix results in a $M \times D$ matrix. Also, two matrices of size $M \times N_1$ and $M \times N_2$ can only be multiplied when $N_1 = N_2$.

D > N

$p > n$ Problem

Consider the following simple situation: You have $N = 1$ and you want to fit $y_1 \approx \beta_0 + \beta_1 x_{11}$, i.e. you want to find β_0 and β_1 given one pair (y_1, x_{11}). Is it possible to find such a line?

This problem is related to something called $p > n$ problem. In our notation, this will be called $D > N$ problem, i.e. the number of parameters exceeds number of data examples.

Similar issues will arise when we use gradient descent or least-squares to fit a linear model. These problems are all solved by using regularization, which we will learn later.