Linear Regression

Machine Learning Context

Supervised

- Given a set of data with associated measurements
- Predict the results of future measurements given a set of known results

Data could be a collection of images, measurements say "this is a duck".

Data could be numerical (such as time intervals) and measurements could be numerical (such as speed of an object or a stock price).

Simplest case is finding a linear relationship.

Basic Problem

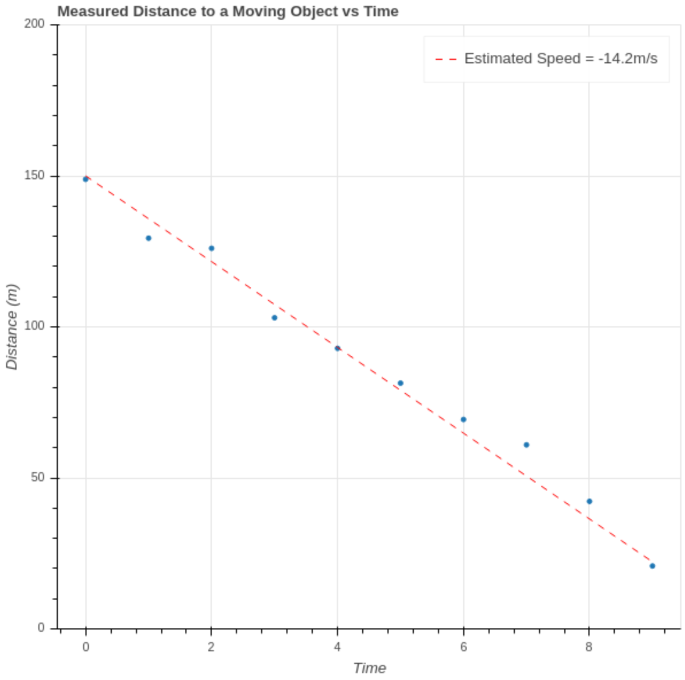


Figure 1: Physics Experiment

Engine size and MPG

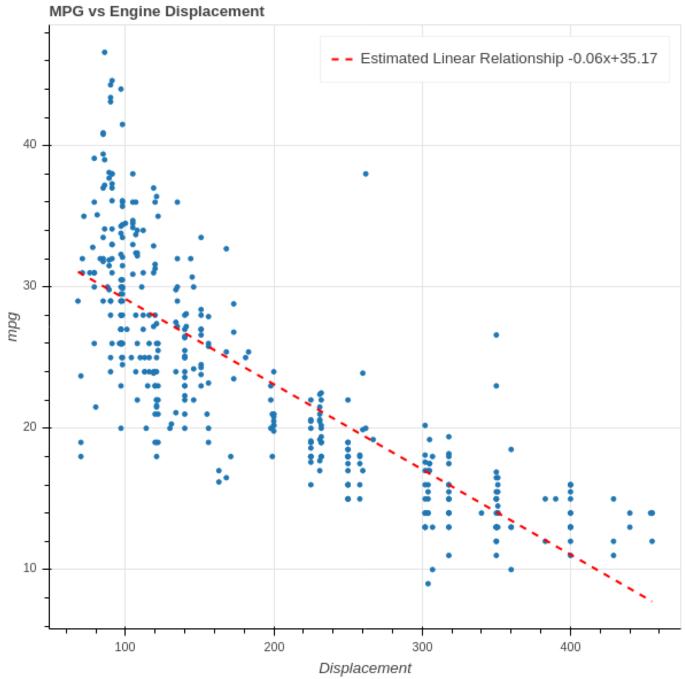
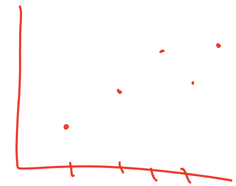


Figure 2: MPG vs Displacement

Mean Squared Error

Data consists of pairs $\{(x_i, y_i)\}$.



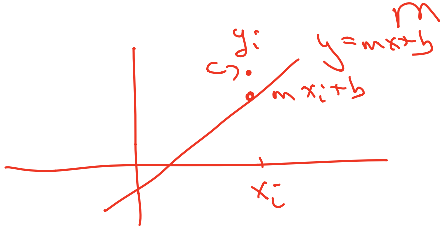
$$MSE(m, b) = \frac{1}{N} \sum_{i=1}^{n} (\underbrace{y_i - mx_i - b}_{\text{difference between observed and predicted value}})^2$$

$Am^2 + Bmb + Cb^2$

Find m and b

$y = mx + b$ minimizing MSE

$mx_i + b$ predicted y value



Minimize MSE

Write E instead of MSE for simplicity.

$$\frac{\partial E}{\partial m} = \frac{1}{N}\sum_{1}^{N} -2x_i(y_i - mx_i - b) = \frac{-2}{N} \sum (x_i y_i - m x_i^2 - b x_i)$$

$$= \frac{-2}{N}\left(\sum x_i y_i - m\sum x_i^2 - b\sum x_i\right)$$

$$S_{xy} - mS_{xx} - b\bar{x}$$

$$\frac{\partial E}{\partial b} = \frac{1}{N}\sum_{1}^{N} -2(y_i - mx_i - b)$$

$$E(m,b) = \frac{1}{N}\sum (y_i - mx_i - b)^2$$

$$\frac{\partial E}{\partial m} = \frac{1}{N}\sum (-2x_i)(y_i - mx_i - b)$$

$$\frac{\partial E}{\partial b} = \frac{1}{N}\sum (-2)(y_i - mx_i - b)$$

$$\frac{\partial E}{\partial m} = -2\left[S_{xy} - mS_{xx} - b\bar{x}\right]$$

$$\frac{\partial E}{\partial b} = -2\left[\frac{1}{N}\sum y_i - \frac{1}{N} m \sum x_i - \frac{b}{N}\sum 1\right]$$

$$= -2(\bar{y} - m\bar{x} - b)$$

Compute the derivatives

$$
\begin{aligned}
\frac{1}{N}\left(\sum_{i=1}^{N} x_i^2\right)m + \quad \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)b &= \quad \frac{1}{N}\sum_{i=1}^{N} x_i y_i \\
\frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)m + \qquad\qquad\qquad\quad b &= \quad \frac{1}{N}\sum_{i=1}^{N} y_i
\end{aligned}
$$

- $\boxed{\overline{x} = \frac{1}{N}\sum x_i}$
- $\boxed{\overline{y} = \frac{1}{N}\sum y_i}$
- S_{xx}, S_{xy}, and S_{yy} are $\frac{1}{N}\sum x_i^2$, $\frac{1}{N}\sum x_i y_i$, $\frac{1}{N}\sum y_i^2$ respectively.

Solve to find the minima

$$\boxed{\begin{aligned} S_{xx} m + \quad \overline{x} b &= S_{xy} \\ \overline{x} m + \quad b &= \overline{y} \end{aligned}}$$

$\frac{\partial E}{\partial m} = 0$

$\frac{\partial E}{\partial b} = 0$

$\overline{x} S_{xx} m + \overline{x}^2 b = \overline{x} S_{xy}$

$\overline{x} S_{xx} m + S_{xx} b = S_{xx} \overline{y}$

$(\overline{x}^2 - S_{xx}) b = \overline{x} S_{xy} - \overline{y} S_{xx}$

$b = \dfrac{\overline{x} S_{xy} - \overline{y} S_{xx}}{\overline{x}^2 - S_{xx}}$

$m =$ similar equation

Solution

$$m = \frac{S_{xy} - \overline{xy}}{S_{xx} - \overline{x}^2}$$

$$b = \frac{S_{xx}\overline{y} - S_{xy}\overline{x}}{S_{xx} - \overline{x}^2}$$

what if $S_{xx} = \overline{x}^2$?

This means: $\frac{1}{N}\sum x_i^2 = \left(\frac{1}{N}\sum x_i\right)^2$

Only happens if all x_i are equal!