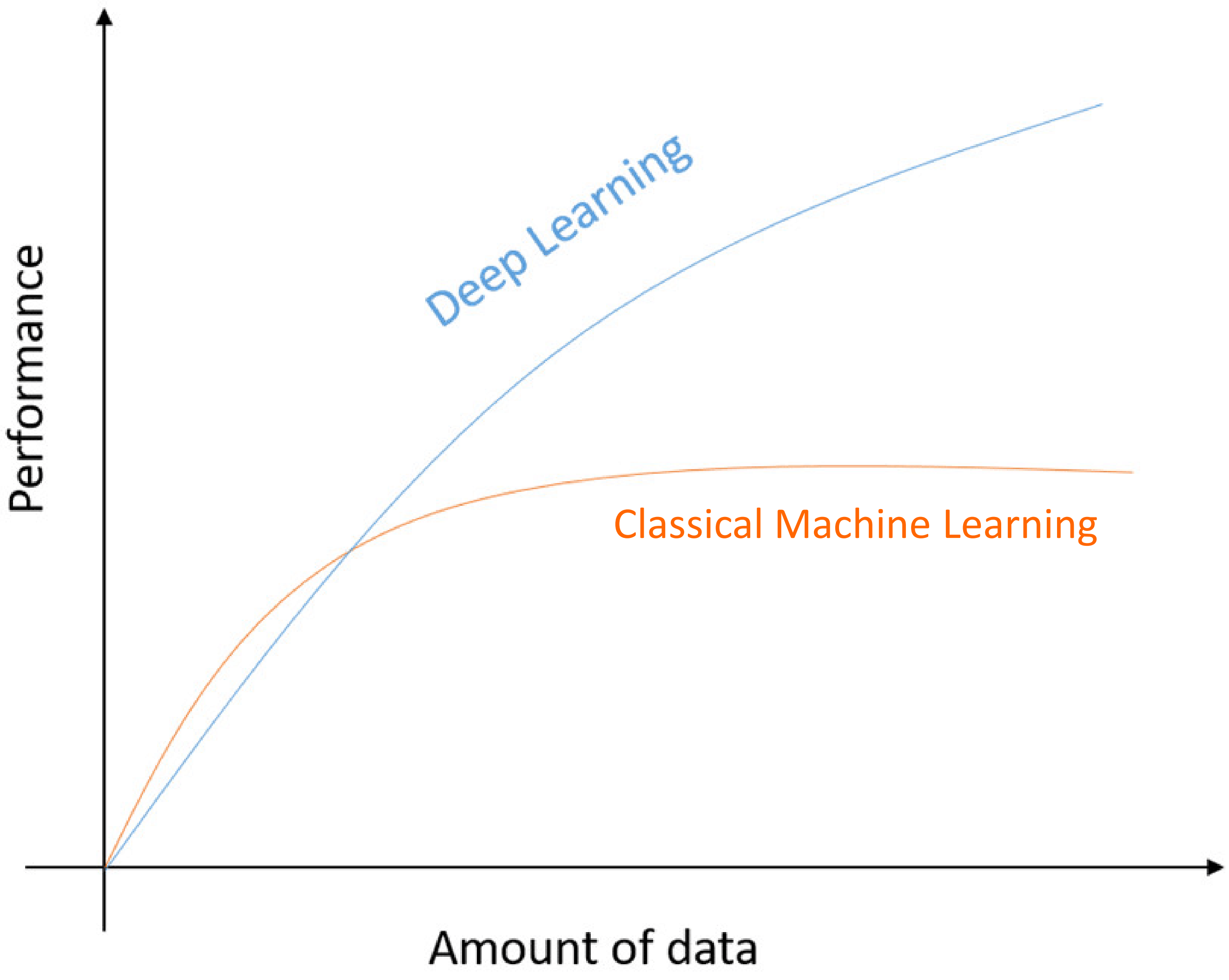
Machine Learning INF2008

Lecture 01: Introduction to Linear Models

Donny Soh

Singapore Institute of Technology

Classical Machine Learning: Linear Regression



Linear Models 1: Linear Regression

Consider an experiment we are familiar with during secondary school.

The photo beside is that of a pendulum.

A period of a pendulum is the time it takes for the pendulum to complete one swing from left to right, to left again.

In our experiments them, we adjusted the length of the pendulum.

We used that to find out how the length of the pendulum affects the period of the pendulum swing.

Linear Models 1: Linear Regression

What are we measuring?

The time it takes for the pendulum to swing from left to right, to left again (also known as the period of a pendulum).

Dependent Variable (y)

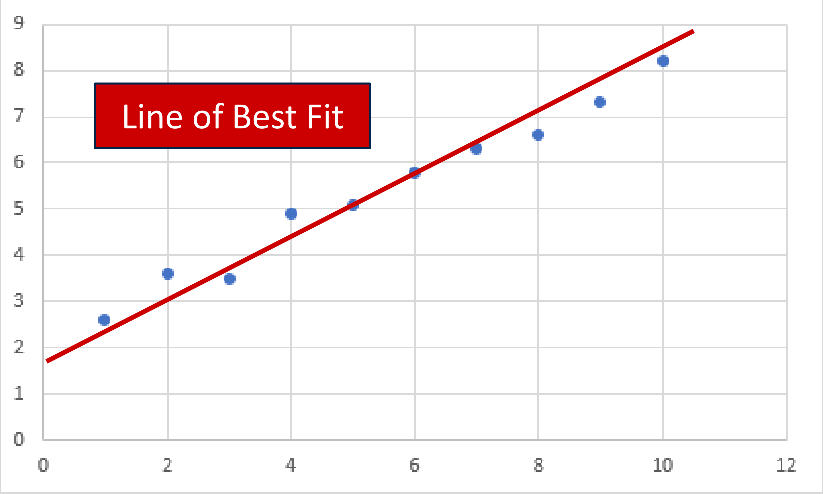
What are we adjusting?

We are adjusting the length of the pendulum to determine how the length of the pendulum affects the period of the pendulum.

Independent Variable (x)

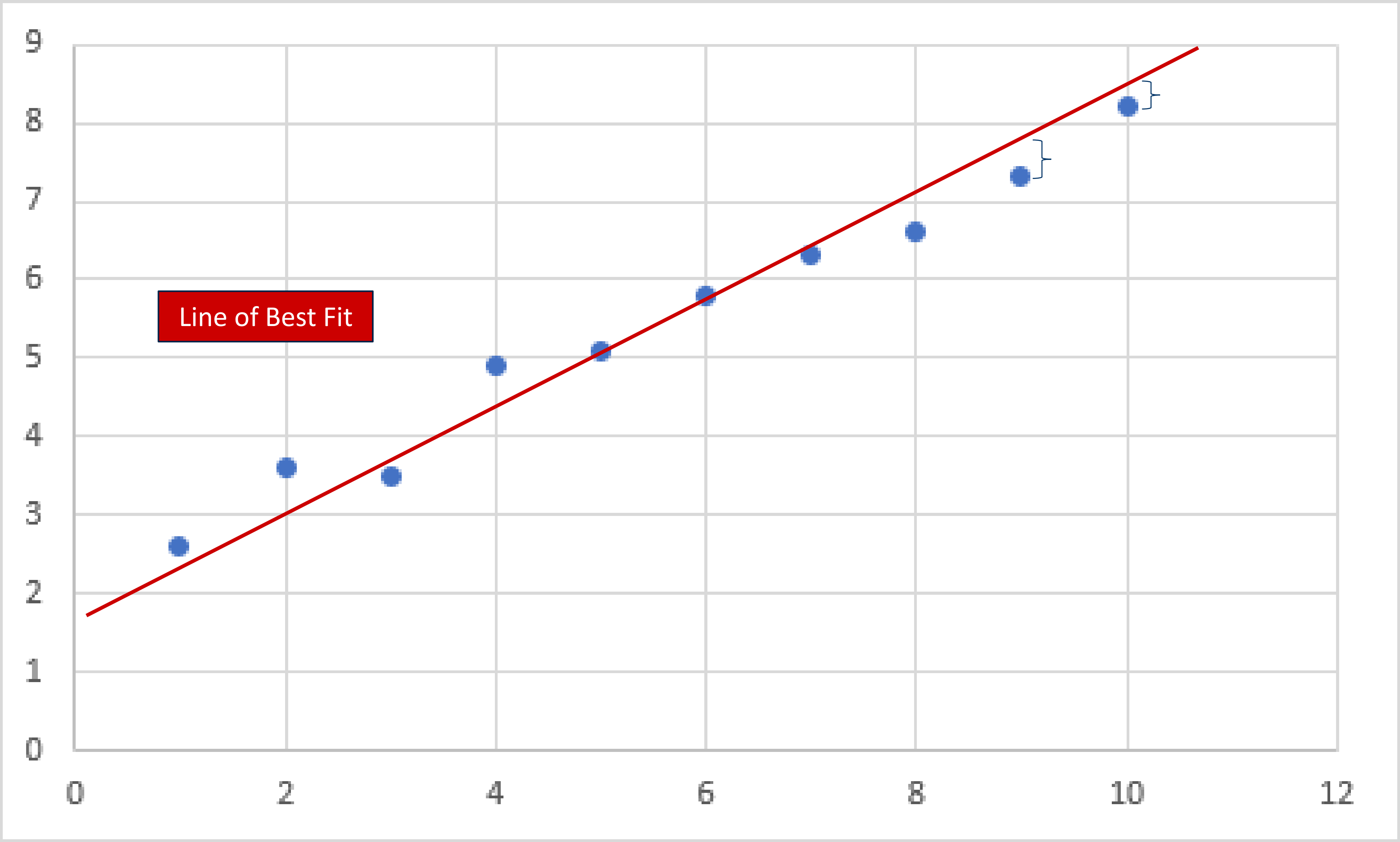
x (cm)	1	2	3	4	5	6	7	8	9	10
y (s)	2.6	3.6	3.5	4.9	5.1	5.8	6.3	6.6	7.3	8.2

$$y = m * x + c$$



Linear Models 1: Linear Regression (Root Mean Square Error)

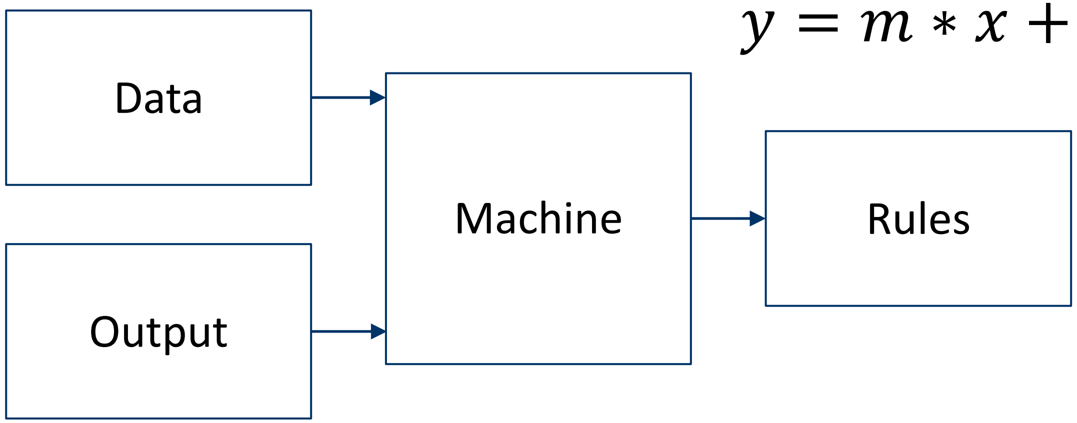
There is a difference from the line of best fit to the actual lines.



$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

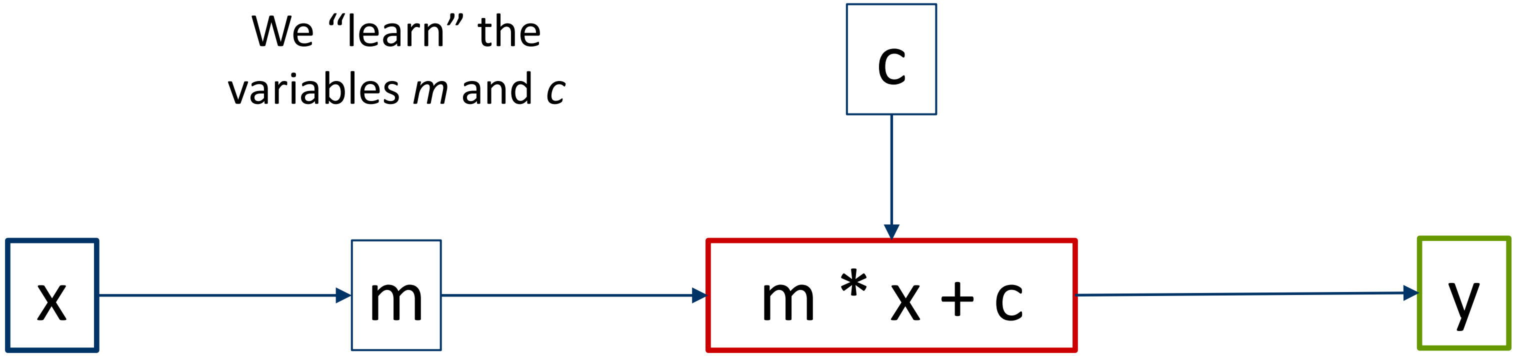
Linear Models 1: Linear Regression

x (cm)	1	2	3	4	5	6	7	8	9	10
y (s)	2.6	3.6	3.5	4.9	5.1	5.8	6.3	6.6	7.3	8.2

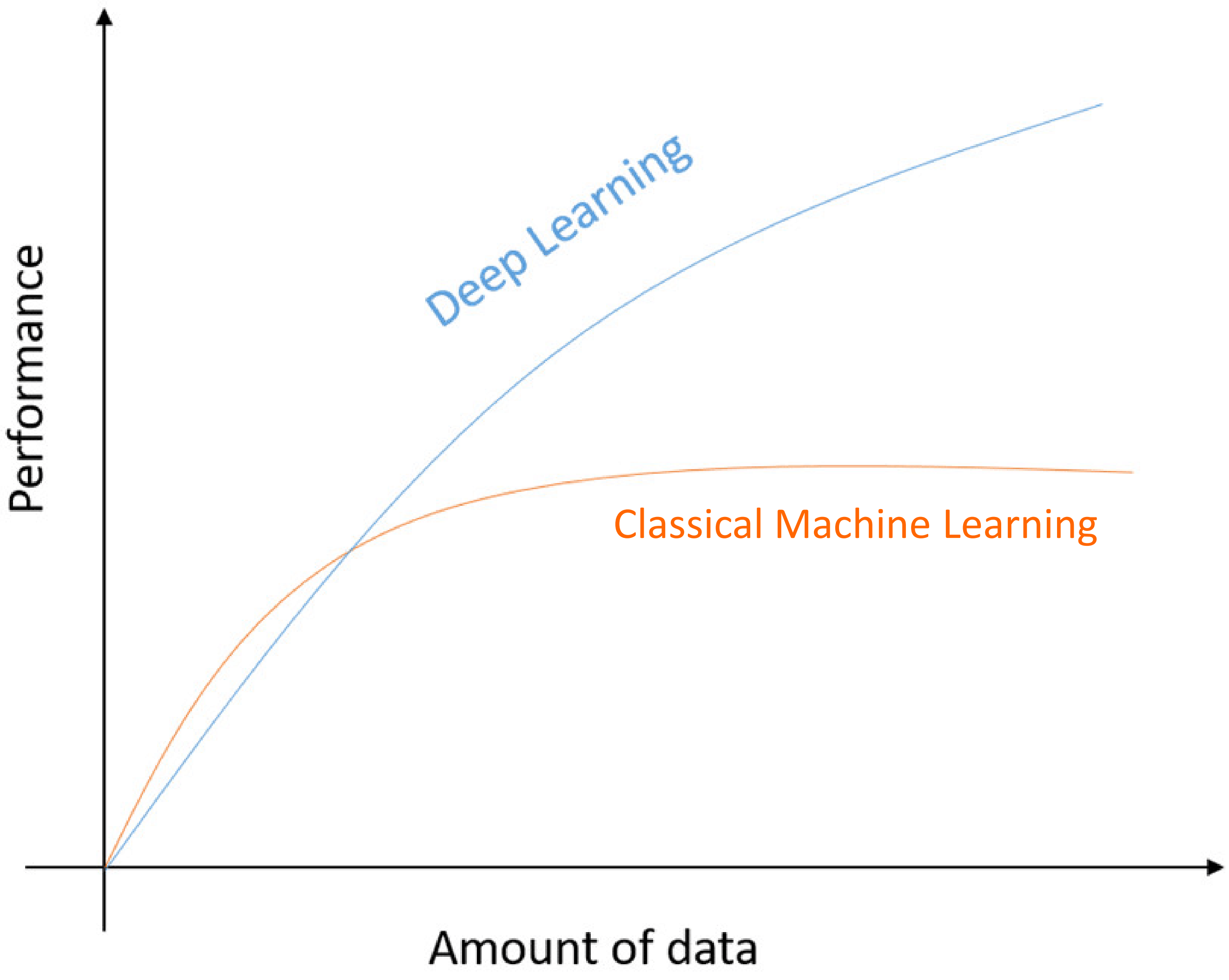


$$y = m * x + c$$

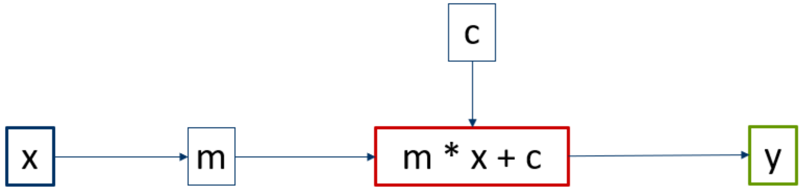
$$y = m * x + c$$



Classical Machine Learning



Linear Regression



Linear Models 2: Multiple Linear Regression

Consider an experiment we are familiar with during secondary school.

The photo beside is that of a pendulum.

A period of a pendulum is the time it takes for the pendulum to complete one swing from left to right, to left again.

In our experiments them, we adjusted the length of the pendulum **and the weight of the pendulum**.

We used that to find out how the length of the pendulum **and the weight of the pendulum** affects the period of the pendulum swing.

Linear Models 2: Multiple Linear Regression

What is the dependent variable?

The time it takes for the pendulum to swing from left to right, to left again (also known as the period of a pendulum).

Dependent Variable (y)

What are the independent variables?

We are adjusting the length of the pendulum and the weight of the pendulum.

Independent Variables (x_k)

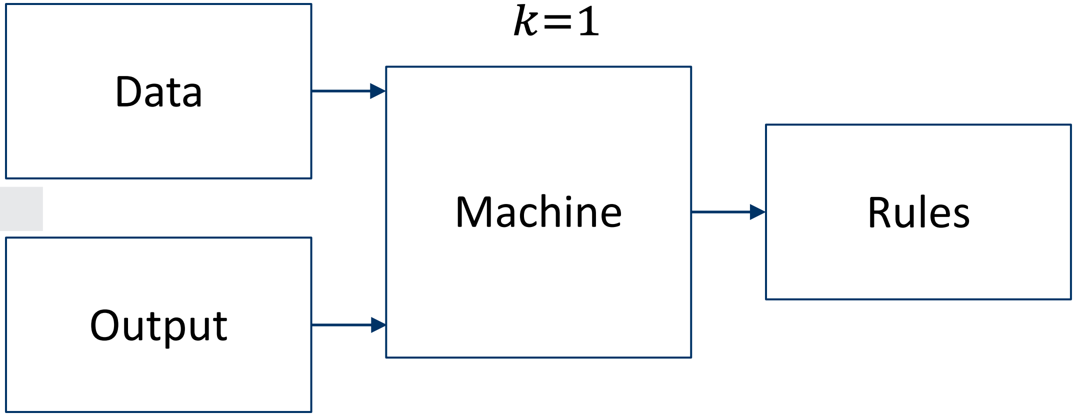
x_1 (cm)	1	1	1	4	4	4	7	7	7	10	10	10
x_2 (g)	10	20	30	10	20	30	10	20	30	10	20	30
y (s)	2.6	2.4	2.5	4.9	4.7	4.5	6.3	6.6	6.5	8.2	8.1	8.3

$$y = m_1 * x_1 + m_2 * x_2 + c \longrightarrow y = \sum_{k=1}^{n} m_k * xk + c$$

Linear Models 2: Multiple Linear Regression

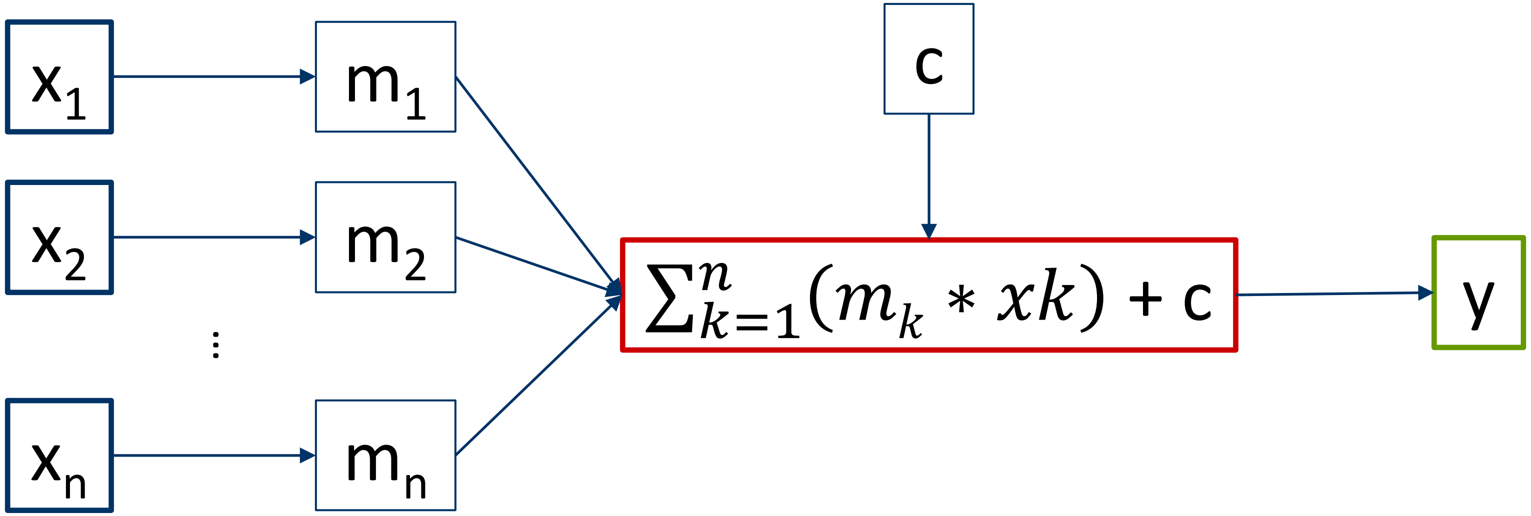
x_1 (cm)	1	1	1	4	4	4	7	7	7	10	10	10
x_2 (g)	10	20	30	10	20	30	10	20	30	10	20	30
y (s)	2.6	2.4	2.5	4.9	4.7	4.5	6.3	6.6	6.5	8.2	8.1	8.3

$$y = \sum_{k=1}^{n} m_k * xk + c$$

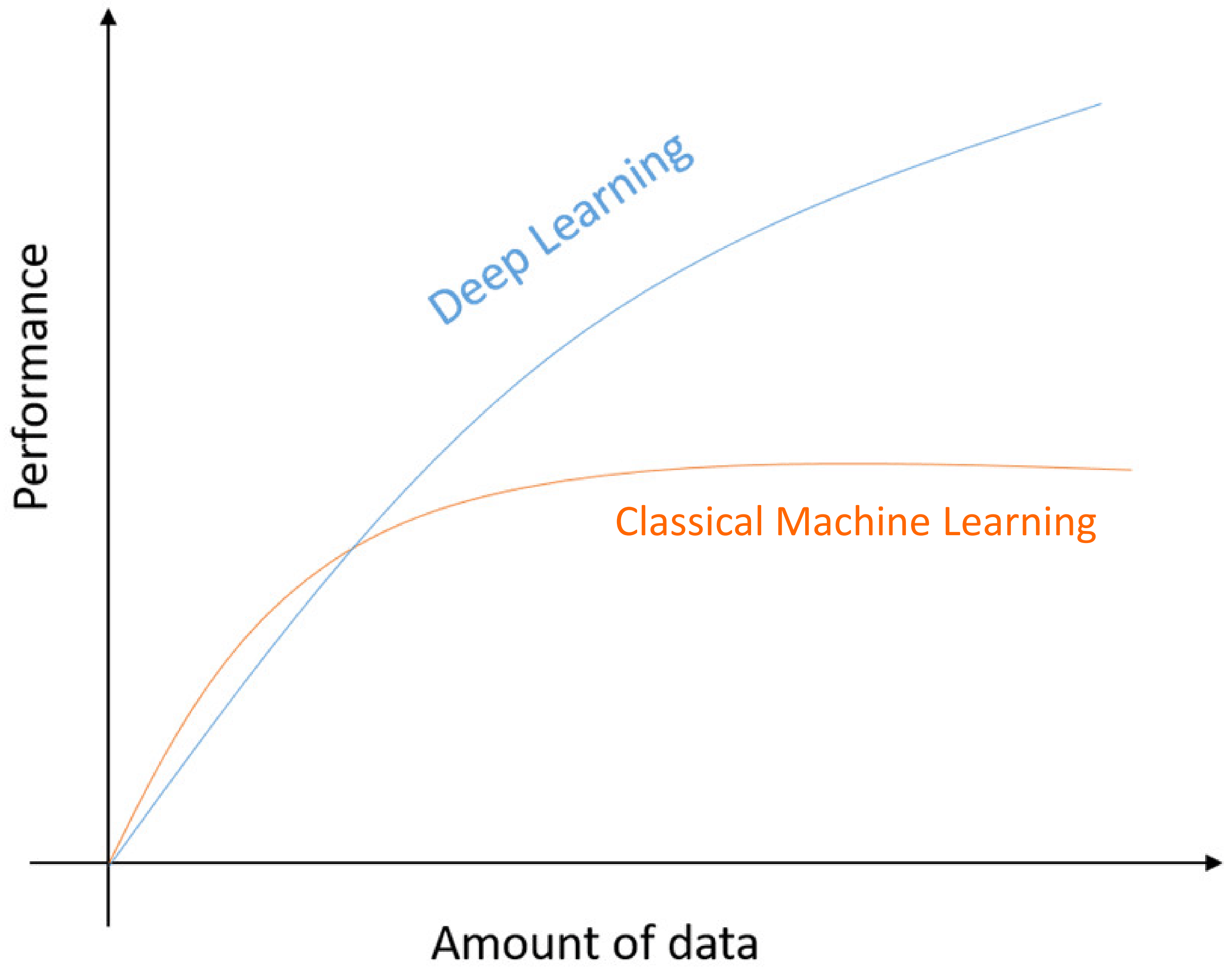


$$y = \sum_{k=1}^{n} m_k * xk + c$$

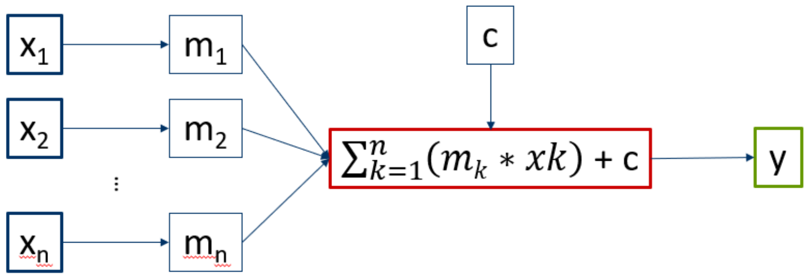
We “learn” the variables m_i and c



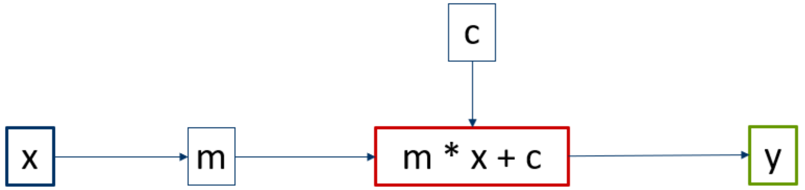
Classical Machine Learning



Multiple Linear Regression



Linear Regression



Fukushima Case Study

Fukushima Case Study

- The safety analysis for the Fukushima Daiichi nuclear power plant was based on historical data dating back to 1600 CE.

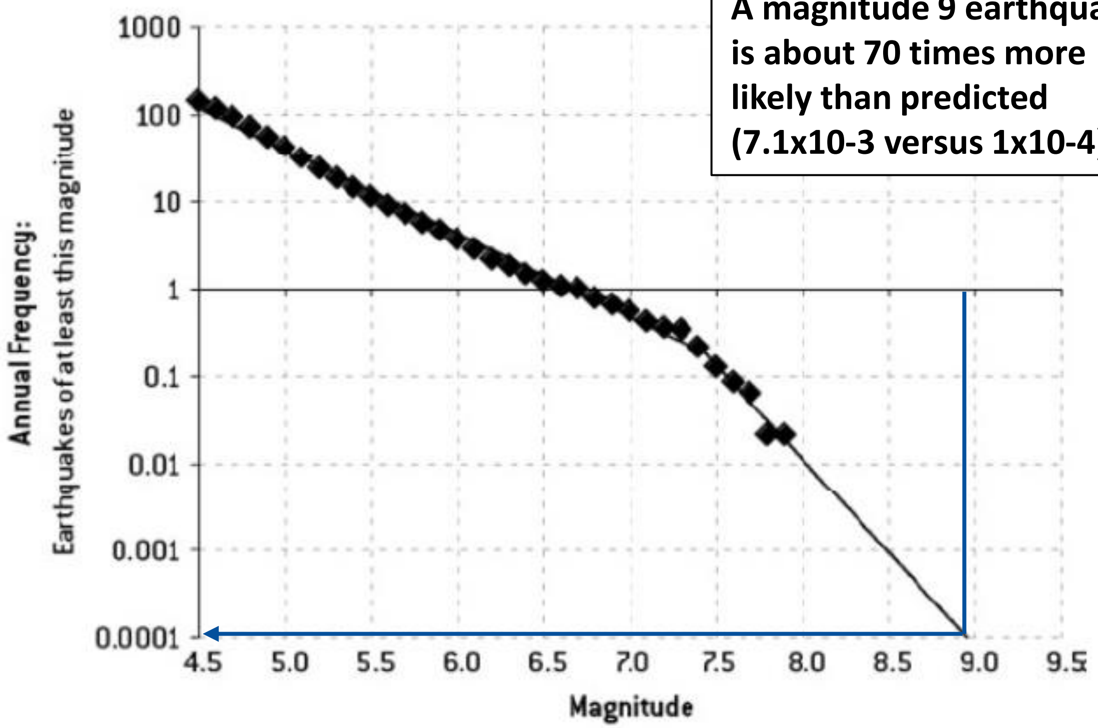
- The plant was designed to withstand a maximum earthquake of 8.6 magnitude, and a tsunami as high as 5.7 meters. The earthquake on March 11, 2011 measured 9.0 and resulted in a >14 meter high tsunami.

- The design basis was developed from a mistake in the regression analysis of the historical earthquake data. The structural engineers responsible for the design overfit their model to the existing data, rather than use the accepted Gutenberg- Richter model, they saw a kink in the data and assumed that the appropriate regression was not linear but polynomial.

Polynomial vs Gutenberg-Richter

Assumed that polynomial regression

FIGURE 5-7C: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
CHARACTERISTIC FIT

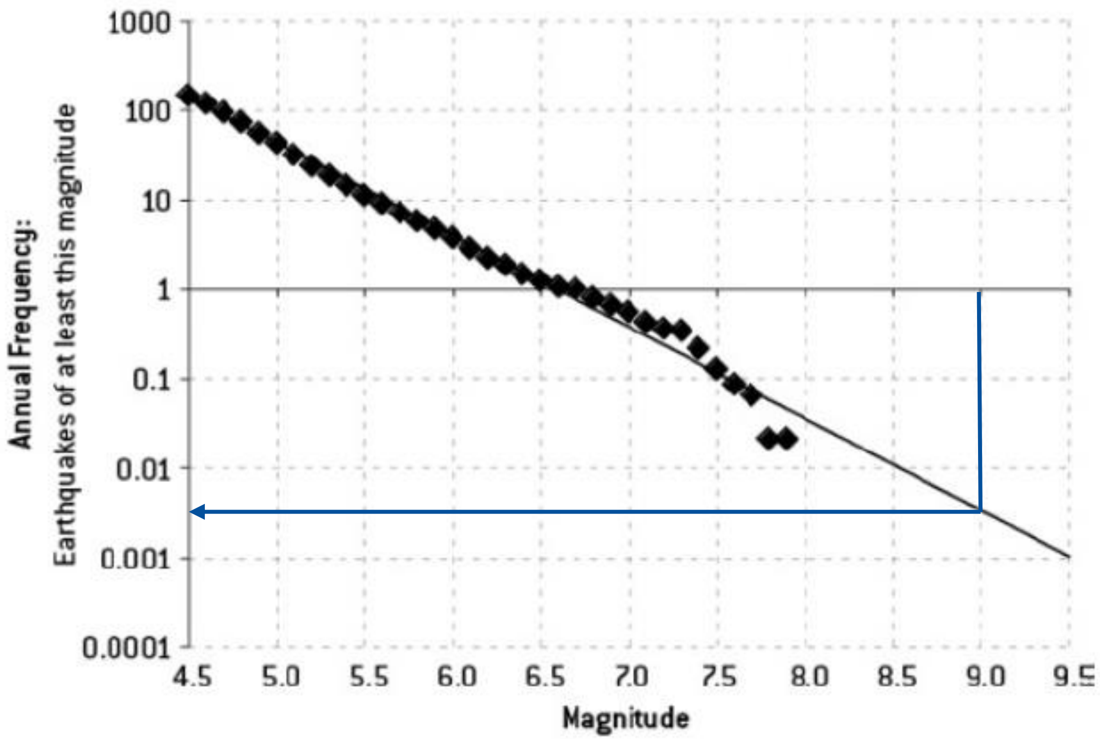


A magnitude 9 earthquake is about 70 times more likely than predicted (7.1x10-3 versus 1x10-4).

(Silver, N, 2012)

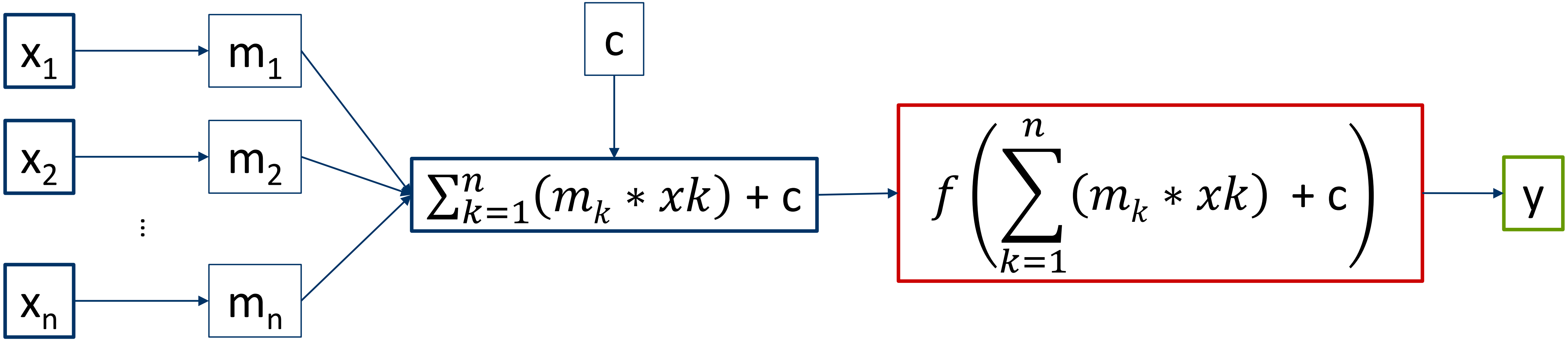
Gutenberg-Richter method uses a simple linear regression

FIGURE 5-7B: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
GUTENBERG-RICHTER FIT



(Silver, N, 2012)

Linear Models 3: Neural Network (Single Layer Perceptron)

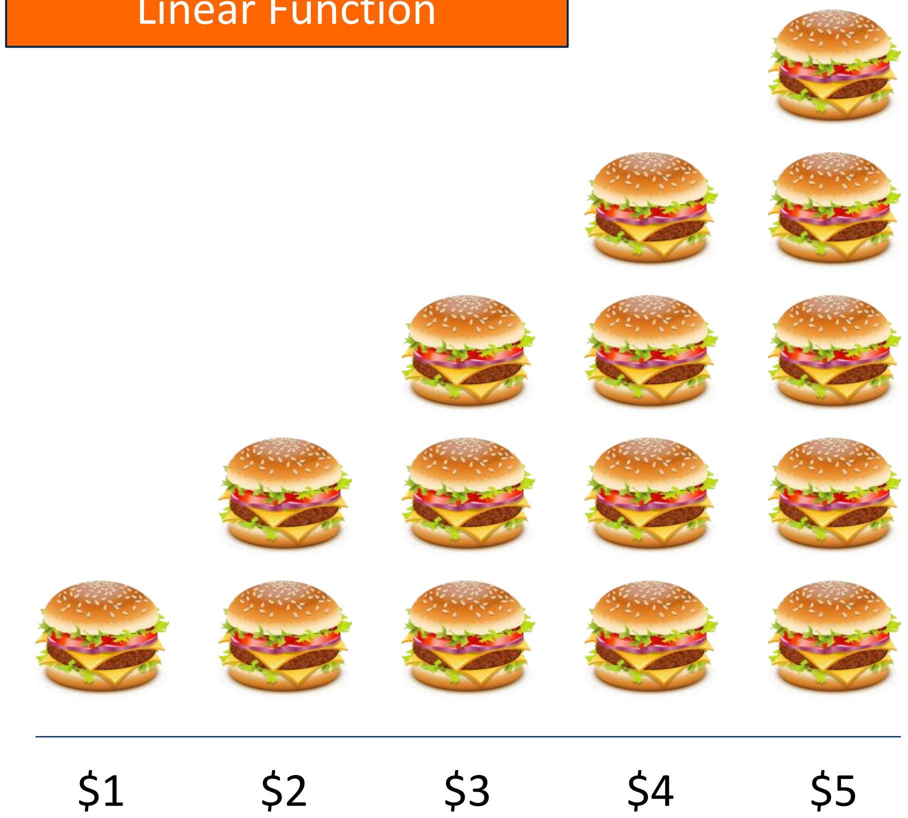


$$y = f\left(\sum_{k=1}^{n} m_k * xk + c\right)$$

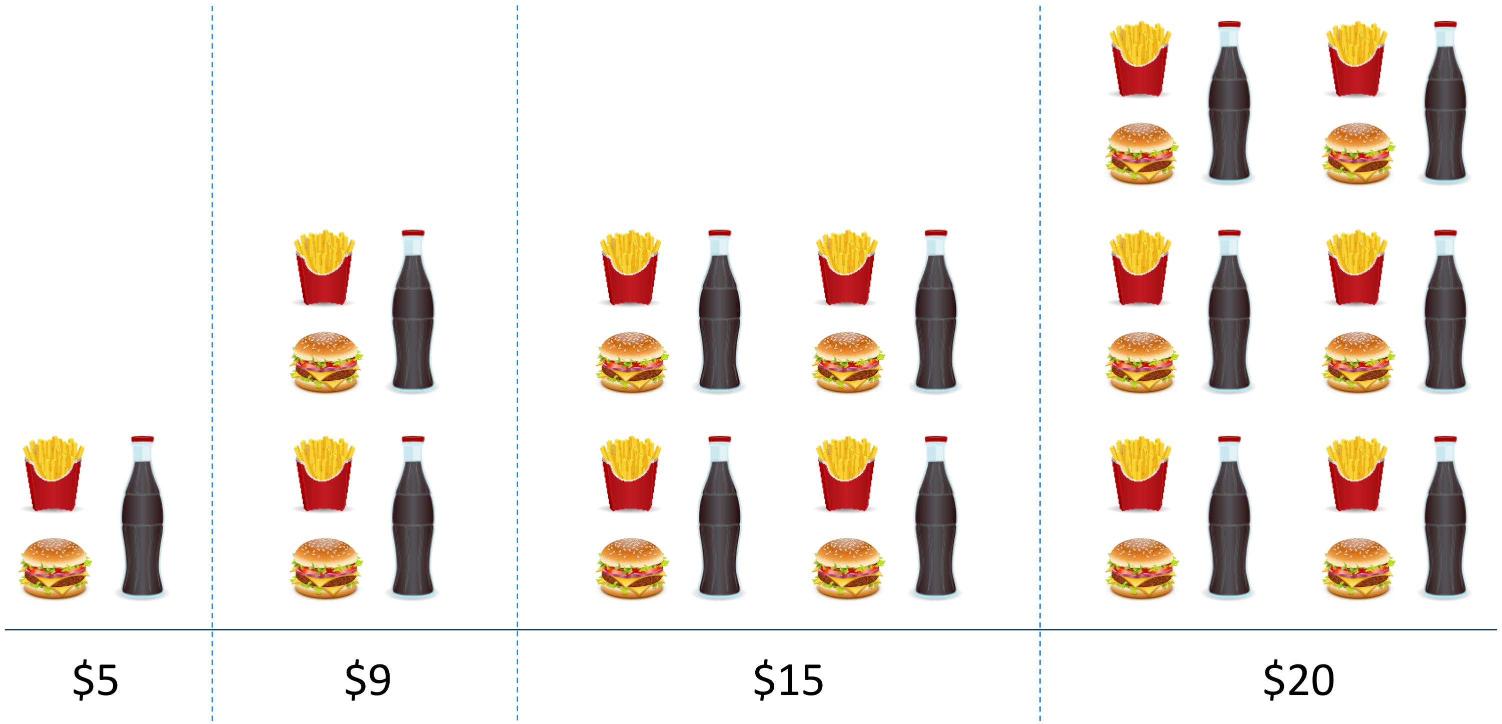
We “learn” the variables m_i and c

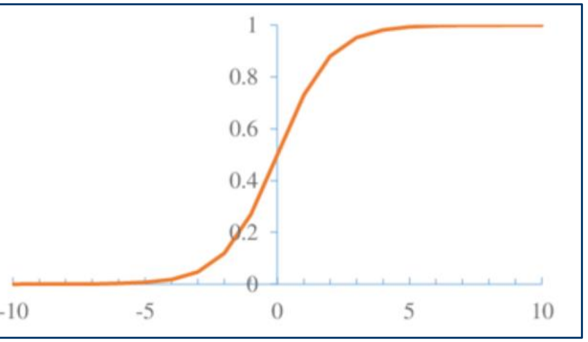
What is a non-linear function?

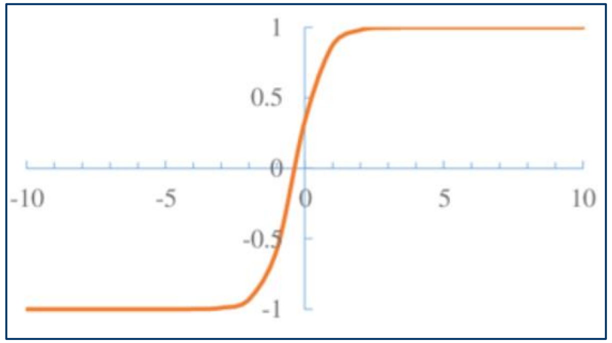
Linear Function

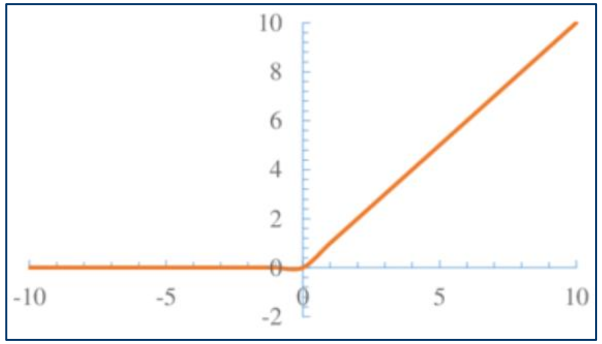


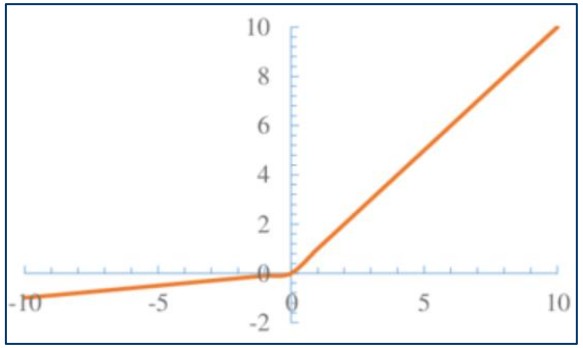
Non-Linear Function



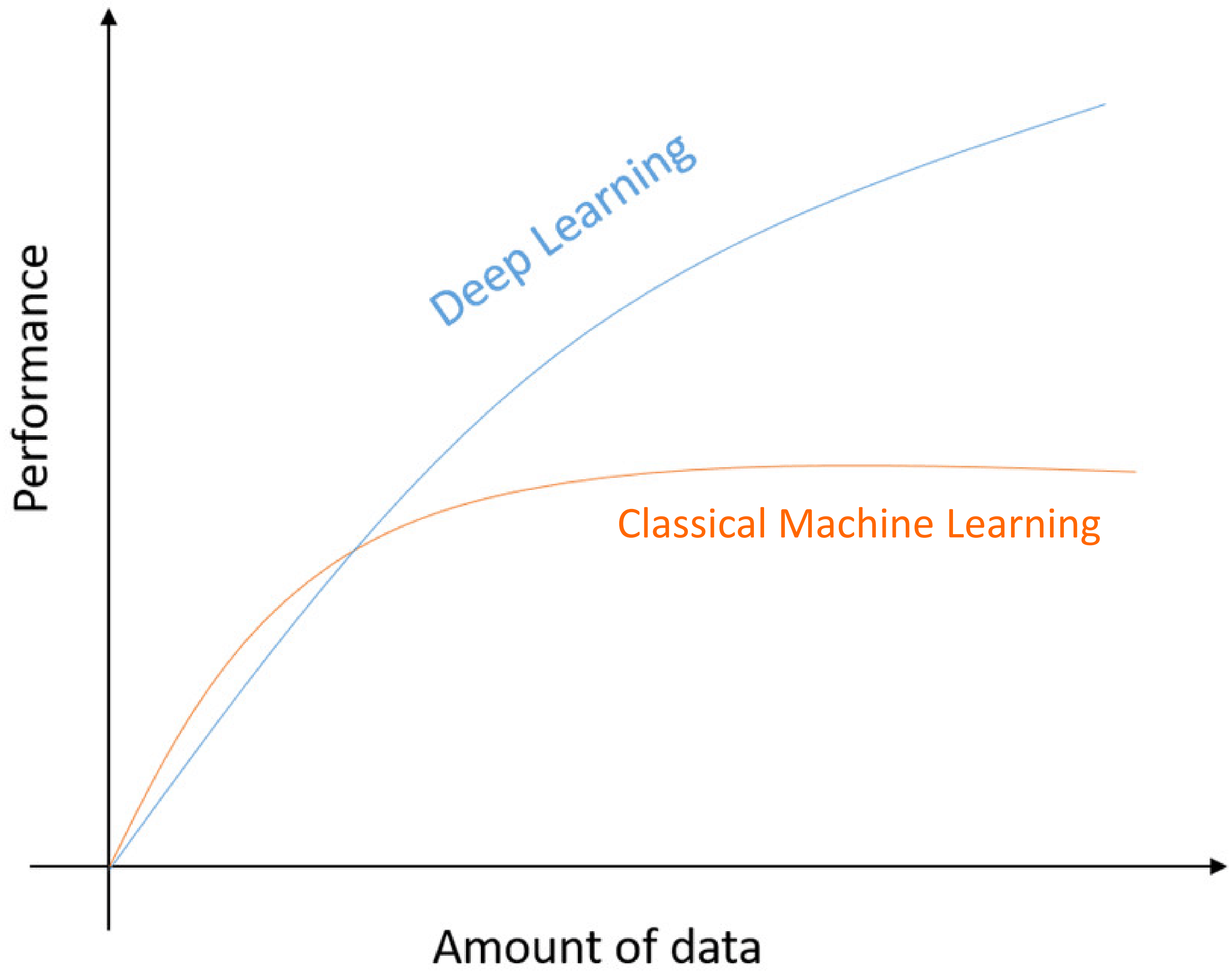




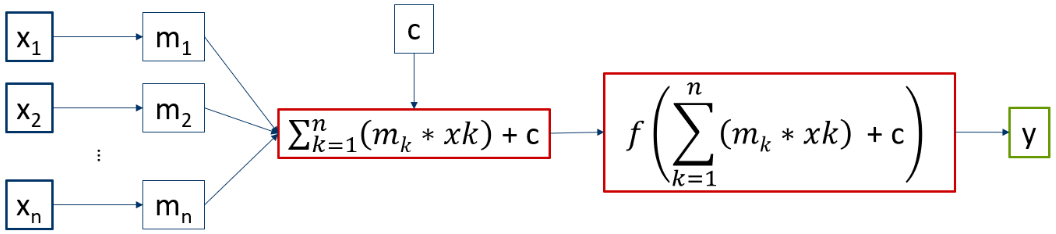




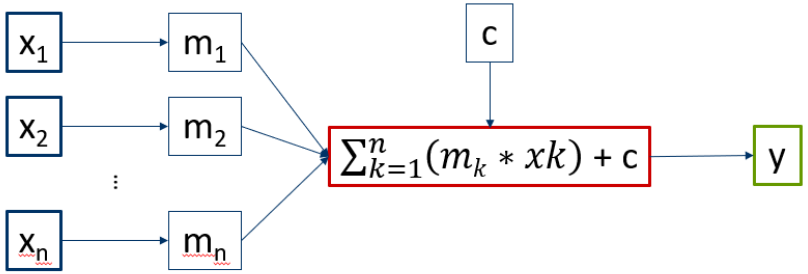
Classical Machine Learning



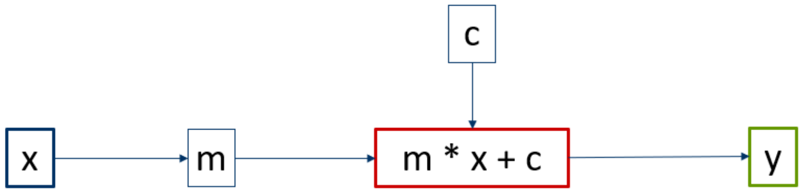
Neural Networks (Single Layer Perceptron)



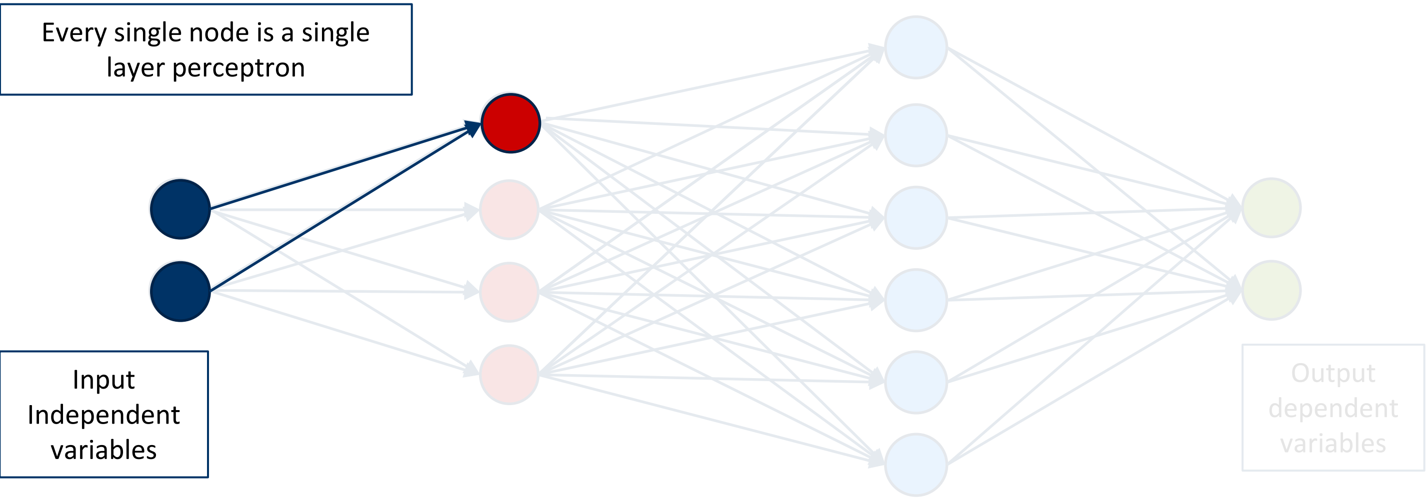
Multiple Linear Regression



Linear Regression



Deep learning versus classical linear models

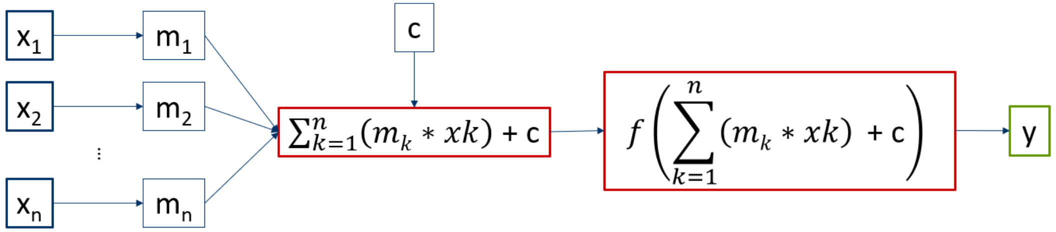


At every node, the neural network **learns** the variables (as in the single layer perceptron)

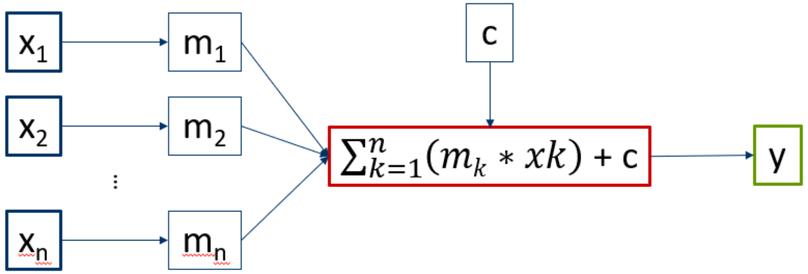
The neural network is **deep** in the sense of having multiple layers.

Hence the term **deep learning**.

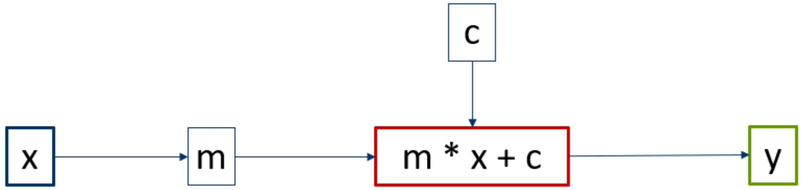
Neural Networks (Single Layer Perceptron)



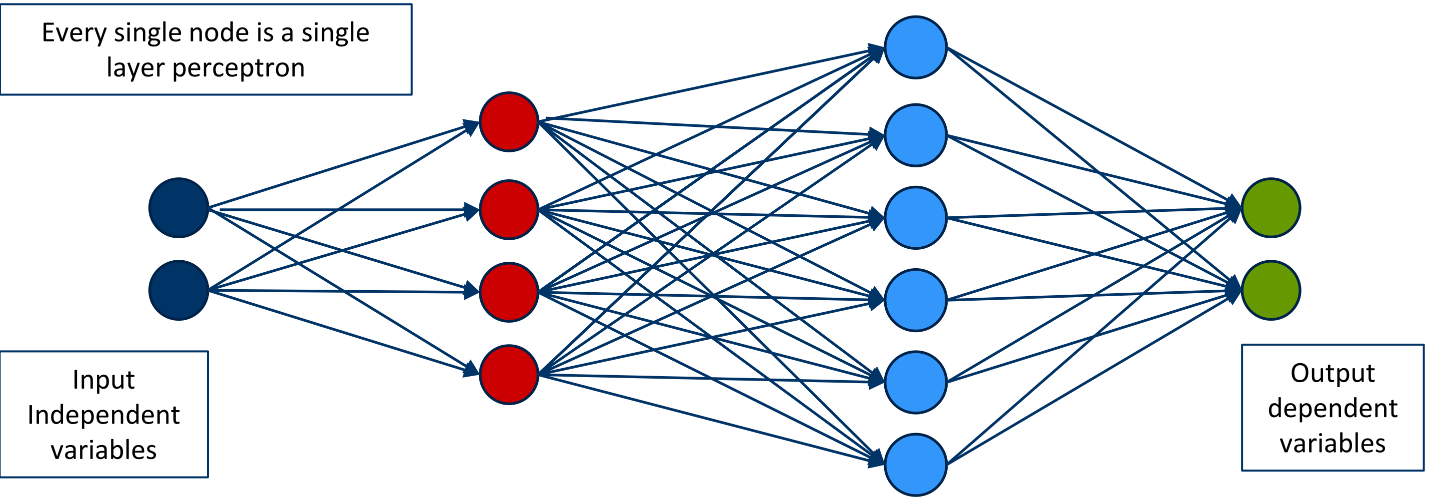
Multiple Linear Regression

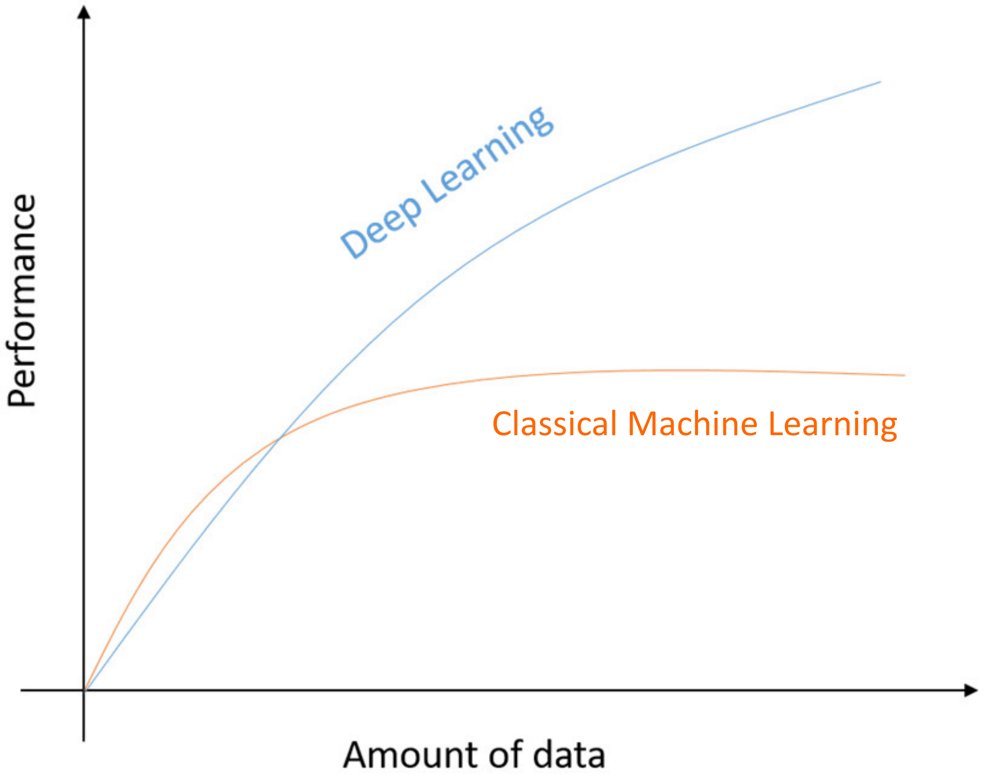


Linear Regression



Deep learning versus classical linear models





At every node, the neural network **learns** the variables (as in the single layer perceptron)

The neural network is **deep** in the sense of having multiple layers.

Hence the term **deep learning**.

Artificial Intelligence, Machine Learning and Deep learning