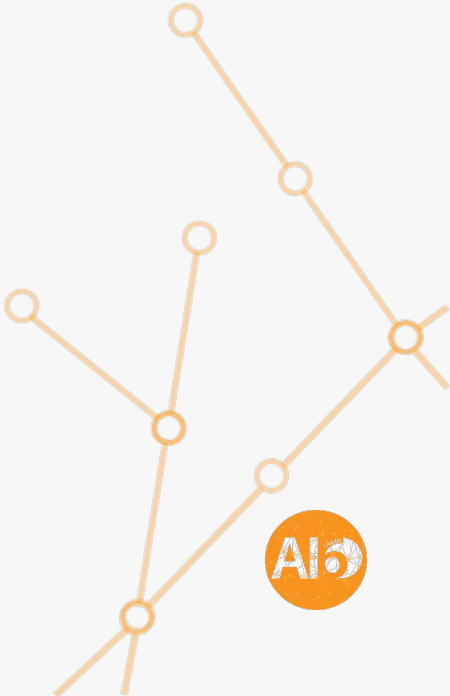
Introduction to Machine Learning

Allen Akinkunle
@allenakinkunle

Agenda

1. Introduction
2. ML Terminology
3. Approaches
4. Walking the Straight Line
5. Linear Regression
6. Squared Loss
7. Training and Test Sets
8. Gradient Descent
9. Overfitting & Regularisation



Introduction

- Machine learning allows computers to acquire knowledge from data without the need for explicit programming.

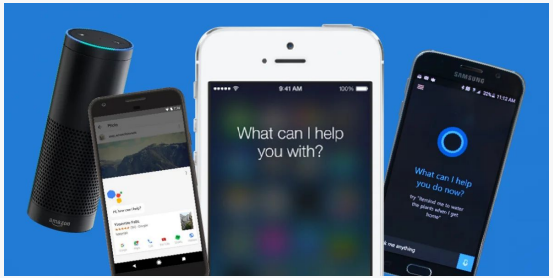
- Machine learning algorithms possess the capability to identify patterns within data and leverage this information to make predictions.

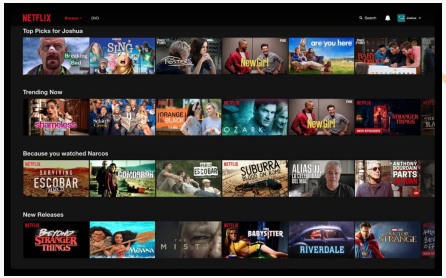
QUESTION?

Why use machine learning?

Applications of ML

- Self-driving cars
- Digital assistants like Siri and Alexa
- Recommendation systems on apps like Netflix
- Fraud detection
- Spam filtering





ML Terminology

- **Label:** A label is the thing we're predicting. It is often denoted by y. Also called dependent/target variable.
- **Feature**: A feature is an input variable that we use for our task. It is often denoted by x. Also called independent variable.
- **Model:** A model encodes the relationship between features and the label.
- **Training:** Training means creating or learning the model.
- **Inference:** Inference means applying the trained model to unseen data.

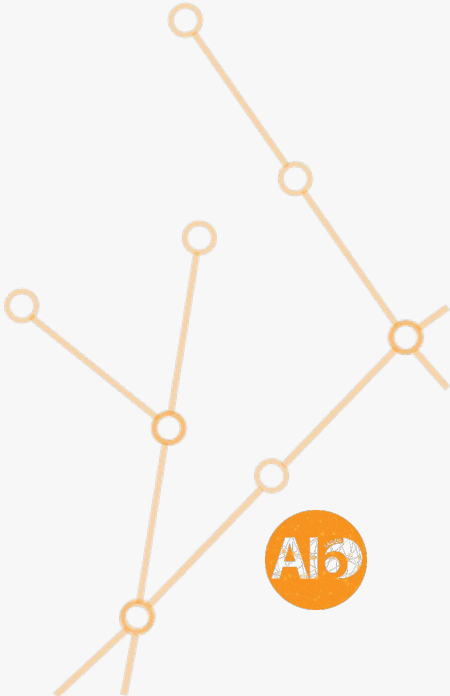
ML Terminology

House Price Prediction

numRooms (feature)	numBaths (feature)	houseValue (label)
2	3	10000
3	1	12500

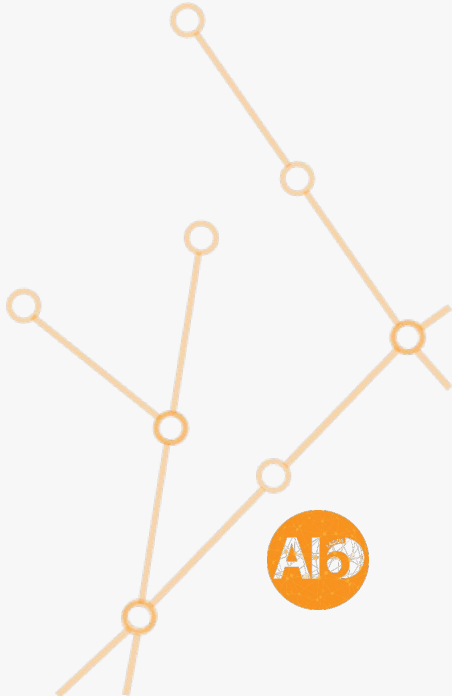
Spam Prediction

Email Body	Spam/No Spam (Label)
Hello Allen, welcome to AI Saturdays	Not Spam
Hello, you have won $100000 in the raffle	Spam



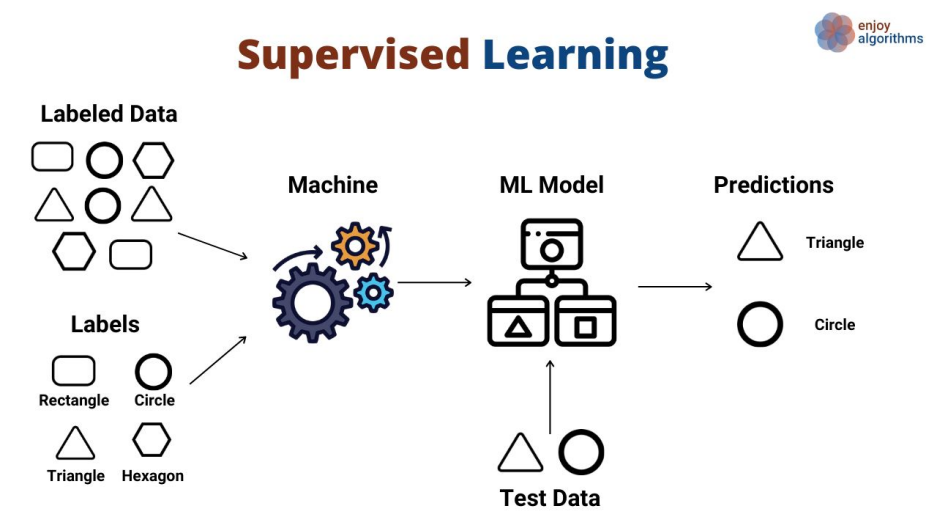
Approaches

- Supervised learning
- Unsupervised learning
- Semi-supervised
- Reinforcement Learning



Supervised Learning

In supervised learning, we use labelled data to train algorithms that are then used for inference.



Supervised Learning

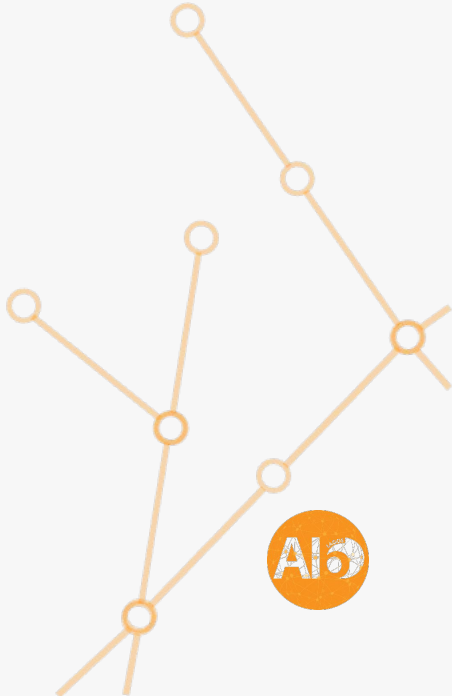
Regression is the task of predicting a continuous quality, while Classification is the task of predicting a discrete class label.

House Price Prediction

numRooms (feature)	numBaths (feature)	houseValue (label)
2	3	10000
3	1	12500

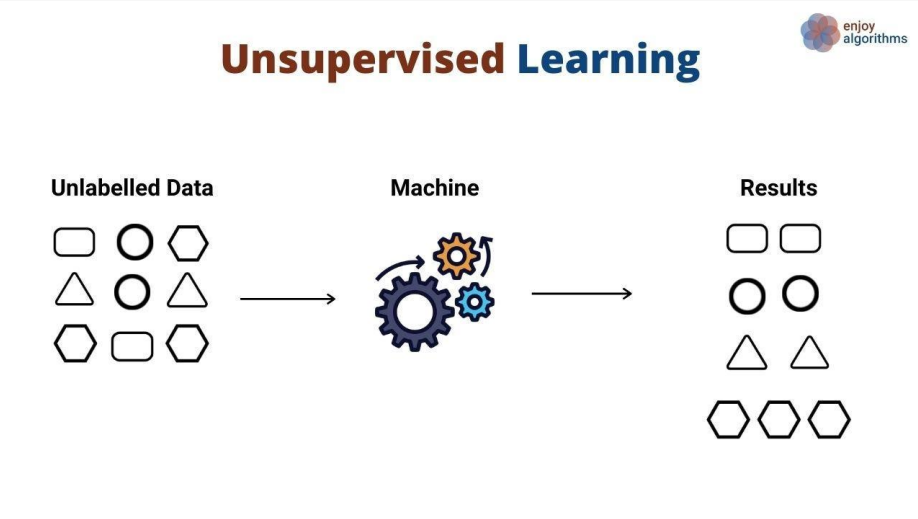
Spam Prediction

Email Body	Spam/No Spam (Label)
Hello Allen, welcome to AI Saturdays	Not Spam
Hello, you have won $100000 in the raffle	Spam



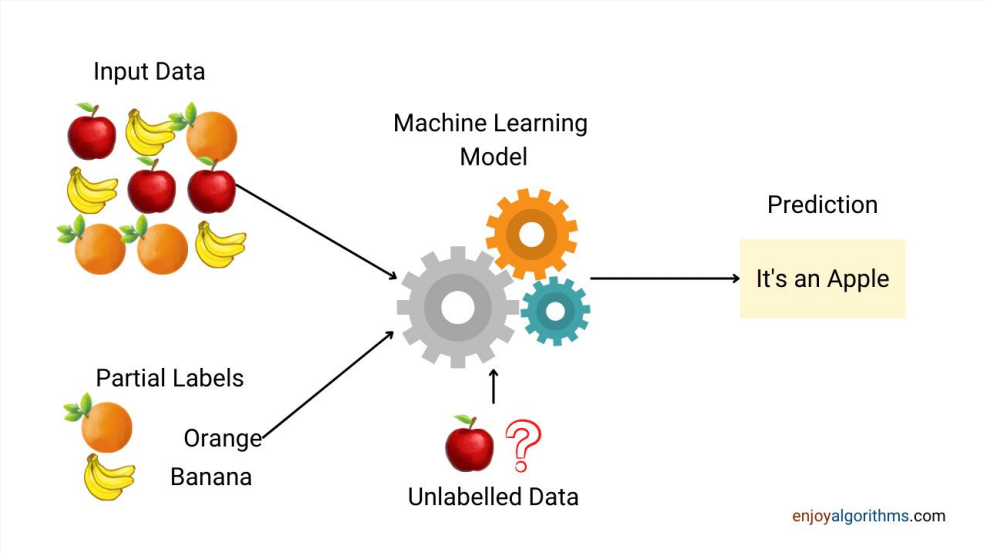
Unsupervised Learning

In unsupervised learning, we use unlabelled data to learn patterns.



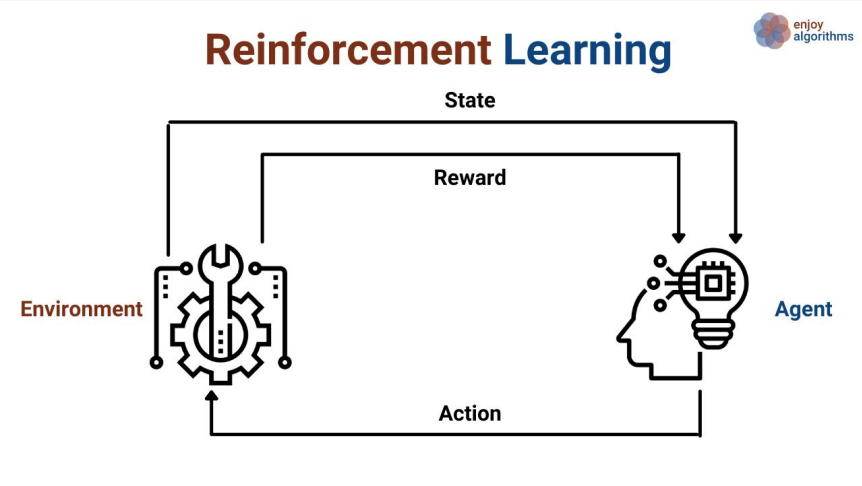
Semi-supervised Learning

Semi-supervised learning falls between supervised learning and unsupervised learning. The dataset contains both labelled and unlabelled data points.



Reinforcement Learning

Reinforcement Learning is based on rewarding desired behaviors and penalising undesired ones. In general, a learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

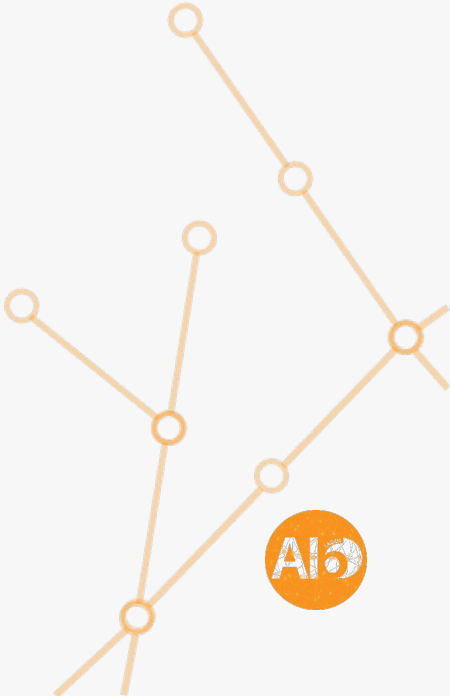


Walking the Straight Line

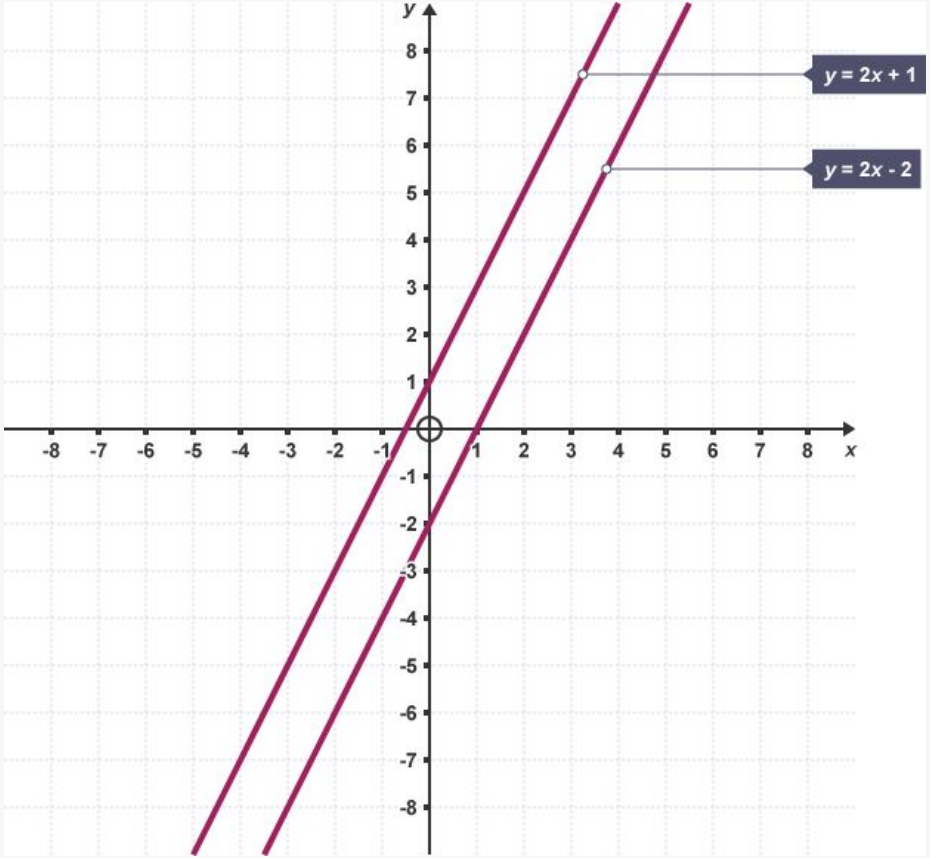
$$y = mx + b$$

Gradient

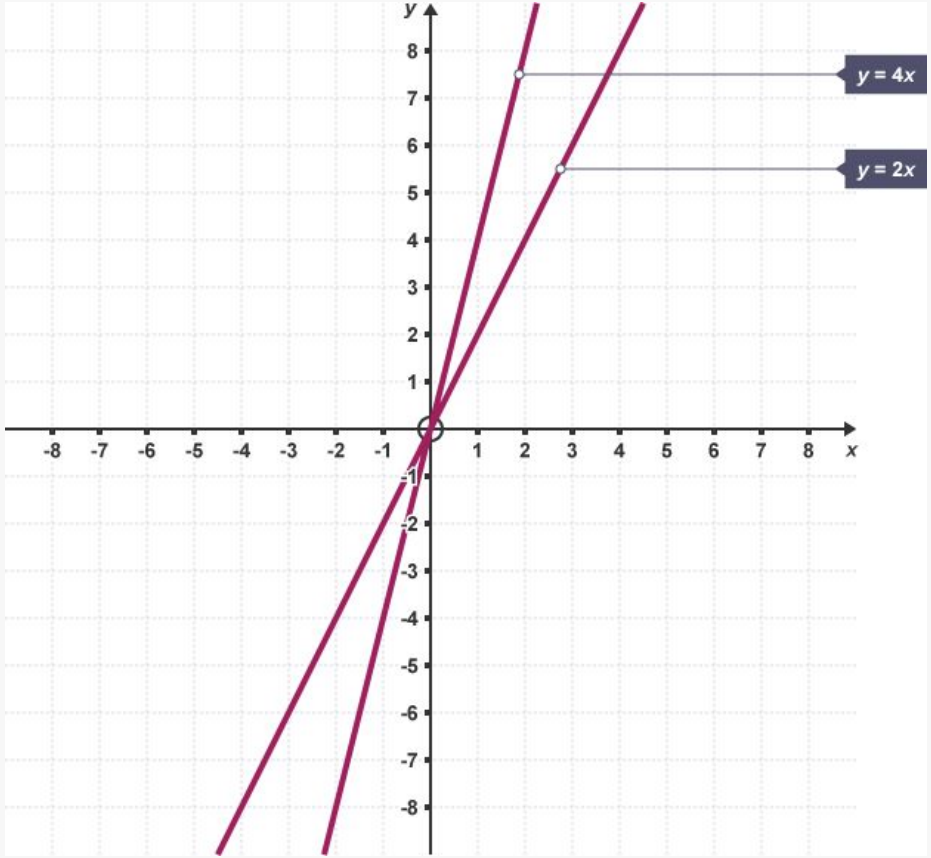
y-intercept

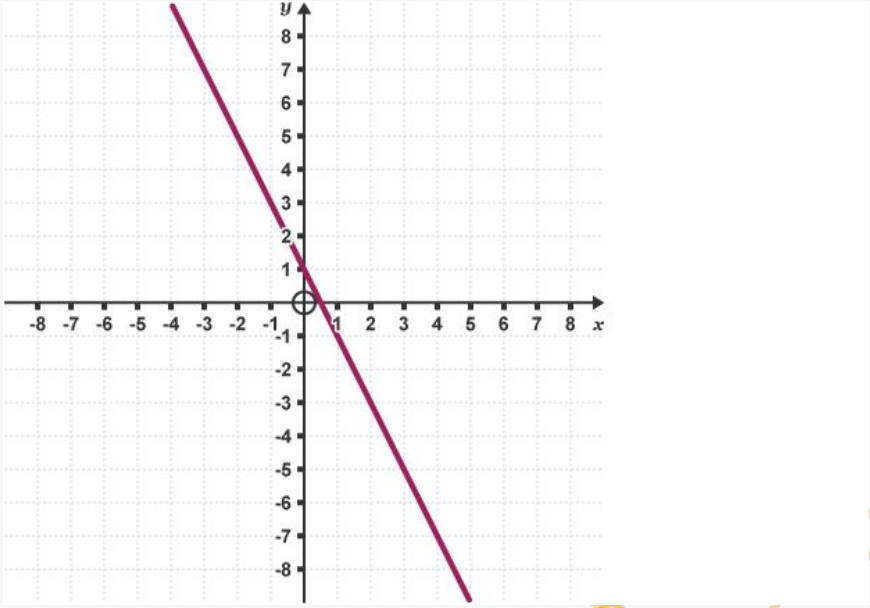


Walking the Straight Line

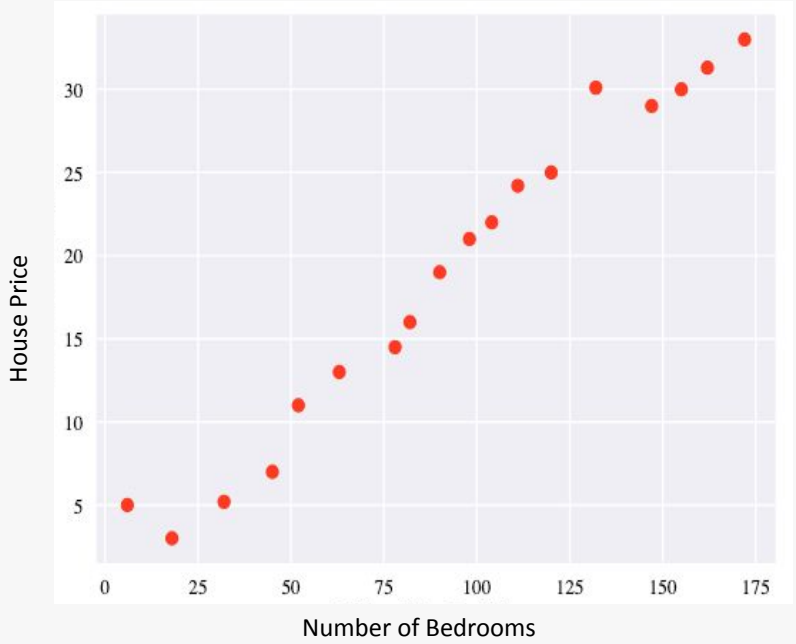


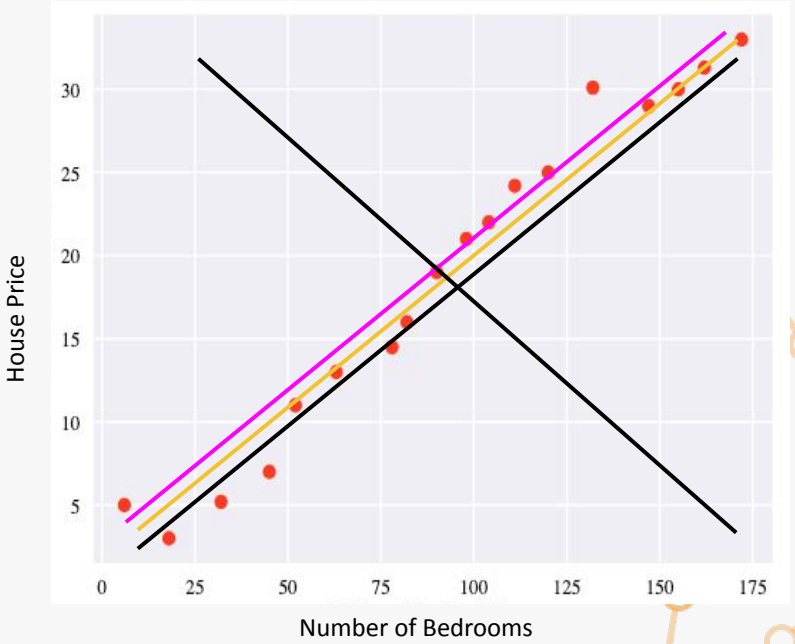
Walking the Straight Line





Linear Regression





Linear Regression

Linear regression assumes that there is a linear relationship between the features and the numerical label.

$$y = mx + c$$

In ML, we usually write the equation as:

$$y = b + w_1 x_1$$

where:

y is the label

b is the bias (y-intercept), sometimes written as w_0

w_1 is the weight of feature x_1

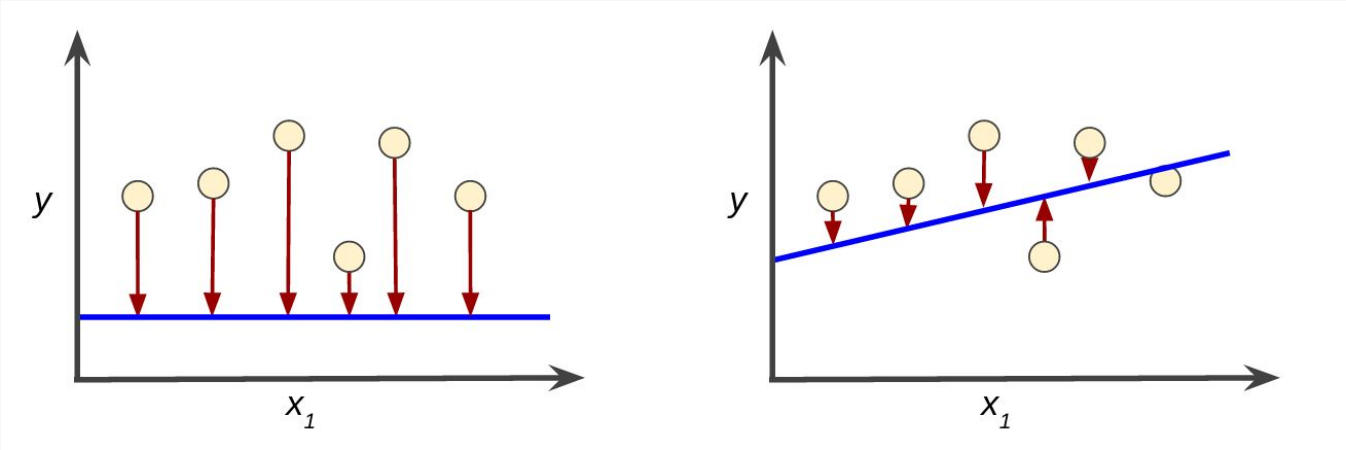
$$\hat{y} = b + w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

Linear Regression

- Training a linear regression model means learning good values for the weights from labelled examples.
- In supervised learning, a machine learning algorithm builds a model that minimises loss. Loss is a penalty for a bad prediction.
- It is a number indicating how bad the model's prediction was on a single data point. If the model's prediction is perfect, the loss is zero, otherwise, the loss is non-zero.

Linear Regression

The goal of training a model is to find a set of weights that have low loss across all the training data points.



Squared Loss

Linear regression uses a loss function called squared loss, and it is the difference between the ground truth and the prediction

Squared Loss

$$= (\text{observation} - \text{prediction})^2$$

$$= (y - \hat{y})^2$$

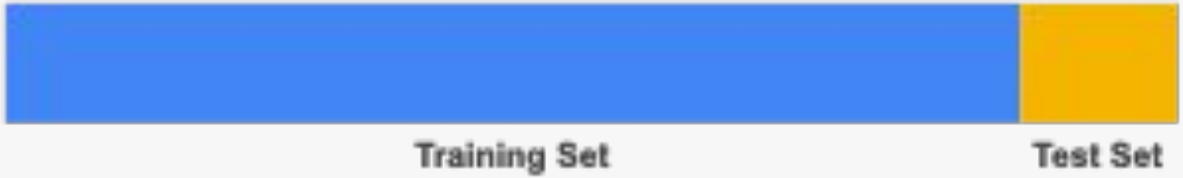
Mean squared error (MSE) is the average squared loss over the whole dataset. Called a cost function.

$$MSE = \frac{1}{N} \sum_{(x,y) \in D} (y - prediction(x))^2$$

Training and Test Sets

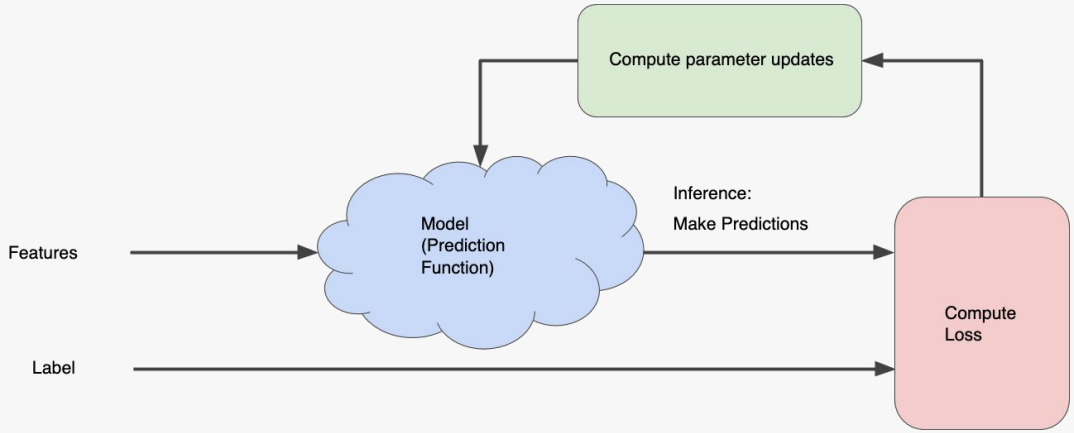
Training set: The algorithm uses this data to learn the underlying patterns and relationship between the dependent and independent variables.

Test set: Kept completely independent from the training data. It is used to assess the model's ability to make accurate predictions on new, unseen data.



Reducing Loss Iteratively

- We begin our training by starting with an initial guess for the weights and biases and iteratively adjust the numbers until we get weights and bias with the lowest loss.



Gradient Descent

Gradient descent is an optimisation algorithm used to minimise our loss function. The idea is to iteratively adjust the weights of our model in the opposite direction of the gradient.

$$w_{i+1} = w_i - \alpha \Delta L(w_i)$$

where:

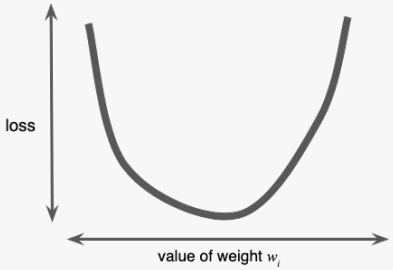
w_i is the current value value of the weight

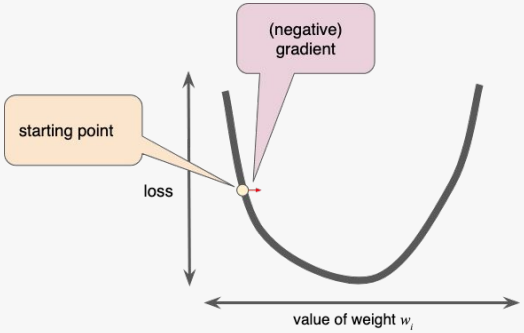
w_{i+1} is the updated value of the weight

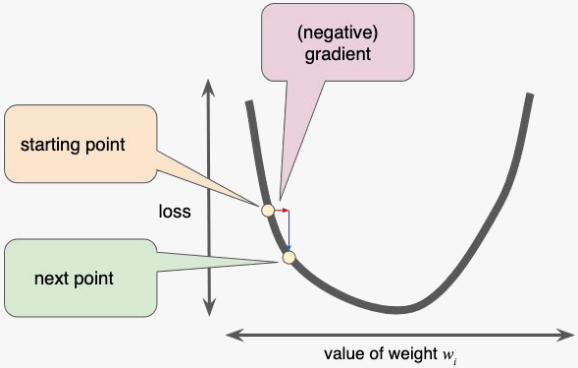
α is the learning rate

$\Delta L(w_i)$ is the gradient of our loss function.

Gradient Descent

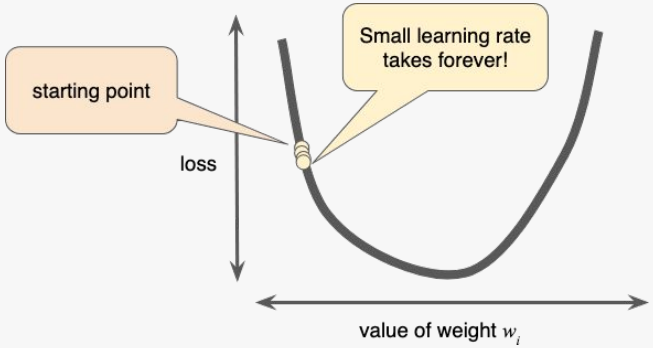




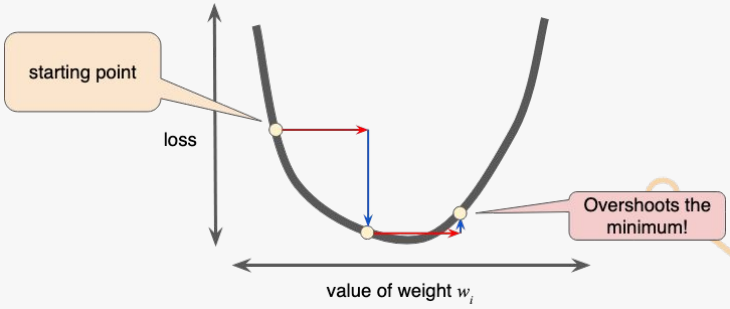


Learning Rate

Learning rate is a hyperparameter that determines the size of the steps taken at each iteration when updating the model's parameters during training.

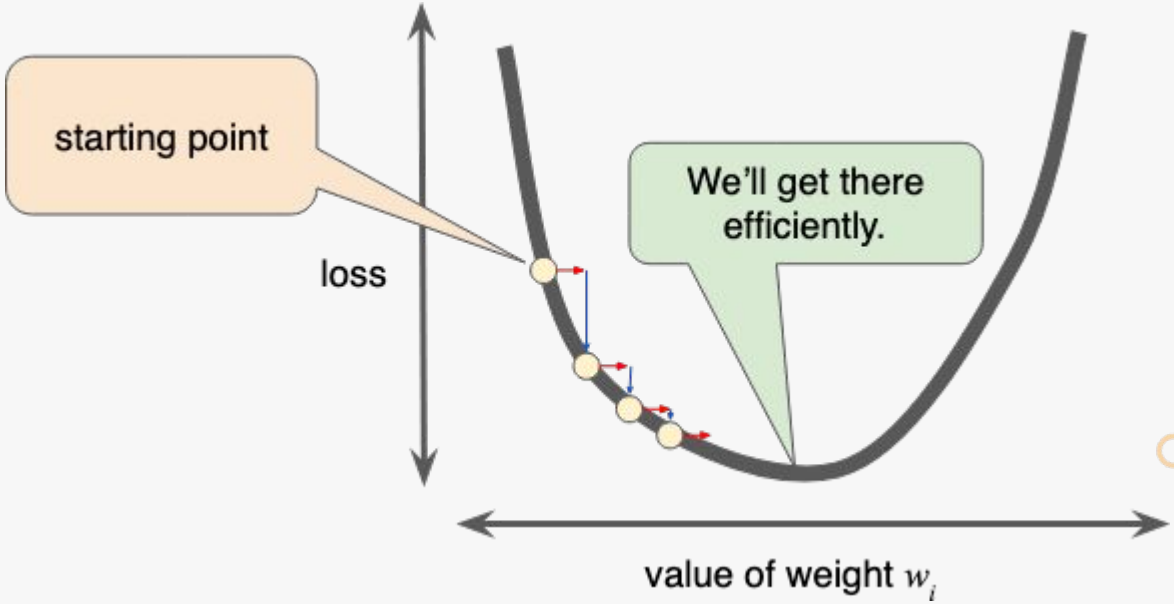


Learning rate too small



Learning rate too large

Learning Rate

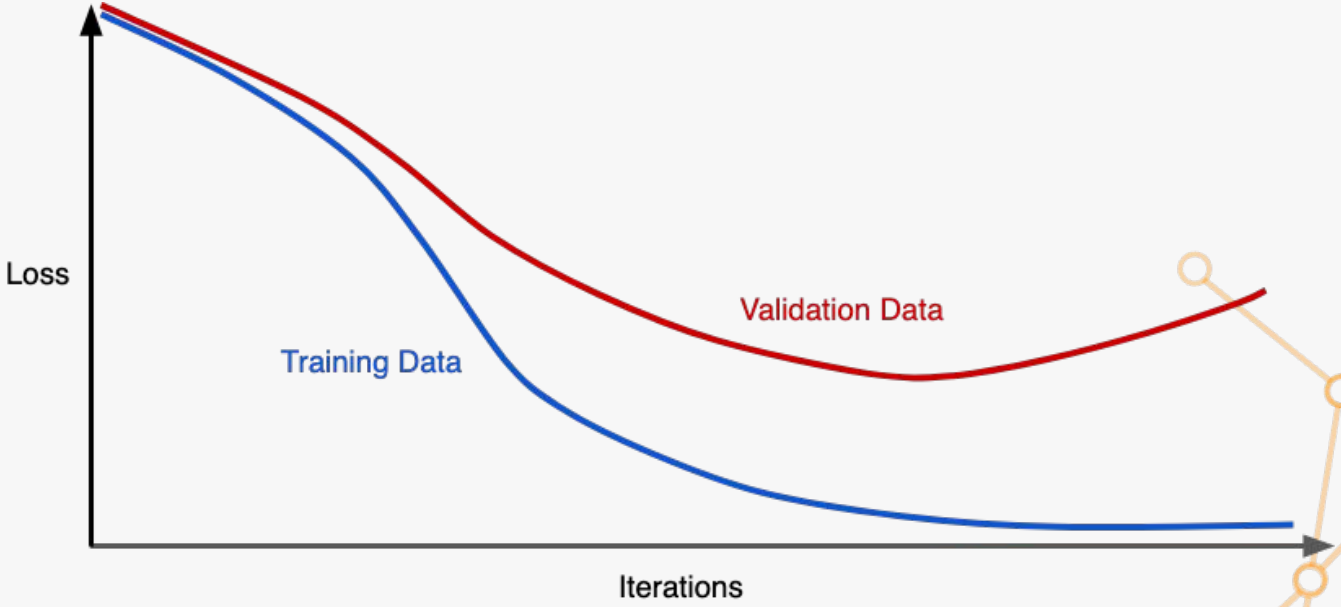


Learning rate just right

Feedback

Overfitting & Regularisation

Overfitting happens when our trained model does not generalise well to new data. It will perform



Thank You