## What is machine learning?

### What is machine learning?

It’s a method of data analysis that automates analytical model building.

### What are neural networks?

A way of modelling biological neurons systems mathematically.

### What is the difference between supervised and unsupervised learning?

Supervised learning use *labelled* examples.

Unsupervised learning uses *unlabelled* examples

### What is the difference between machine learning, neural networks and deep learning?

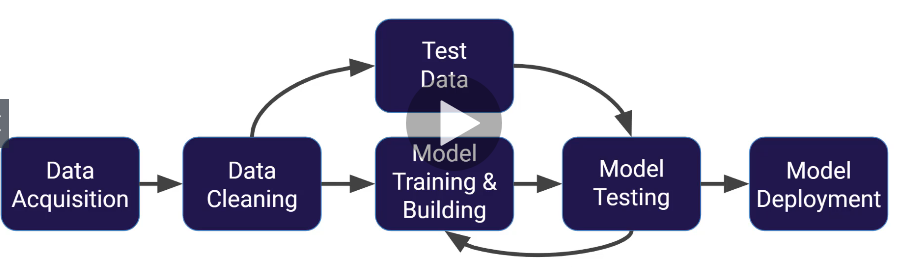
Machine learning – automated analytical models

Neural networks – type of machine learning architecture modelled after biological neurons

Deep learning – A neural network with more than on hidden layer.

## Supervised Learning Overview

### What a flow diagram of the supervised machine learning process?



### Is data split into 2 or 3 sets?

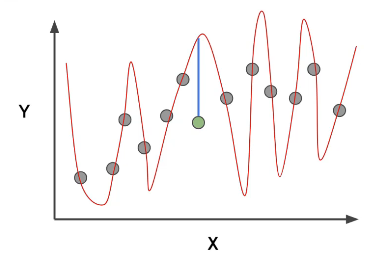
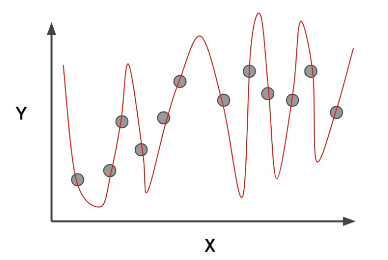
Usually, the data is split into 2 sets. A training set and a test set

Though the best practice is do 3 splits. A training, test, and final validation set

## Overfitting

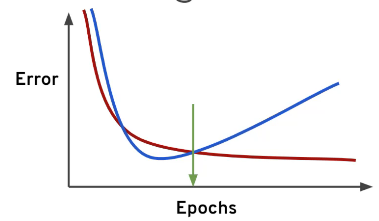
### What is overfitting?

Model fits too much to the noise of the data. Then when comparing with the test data there will be big difference between the training data and the test data.



### How to see if you’re overfitting?

Training data is red and test data is blue. The error will increase on the test data with the model.



## Evaluating Performance – Classification Error Metrics

### What are the key classification metrics?

* Accuracy
* Recall
* Precision
* F1-Score

### Can a single metric determine how good your model is in real life?

No.

### What is Accuracy?

No. of correct predictions/total number of predictions

### When is it good to use Accuracy?

When the data set is balanced in terms of the classes.

### What is Recall and Precision?

They are metrics used for unbalanced classes.

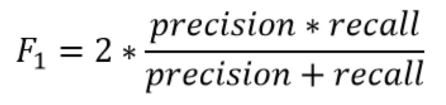
Recall is true positives/(true positives+false negatives).

Recall is true positives/(true positives+false positives).

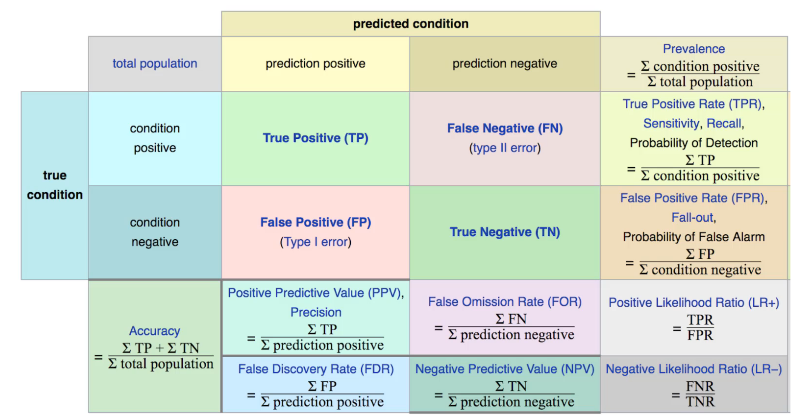
They are complimentary to one another, so you have a trade-off between Recall and Precision.

### What is the F1 Score?

A harmonic mean combination of Recall and Precision, which punishes incorrect data very well.



### What is the confusion matrix?



### What is a good enough accuracy?

This all depends on the context of the situation

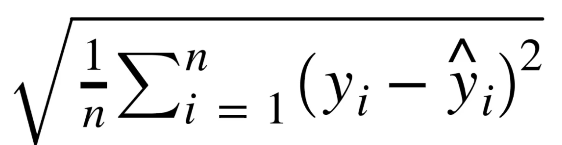
## Evaluating Performance – Regression Error Metrics

### Can you use the metrics from Categorical metrics?

No. They don’t work for regression models.

### What is the best metric for Regression models?

Root mean squared error.



### What’s the advantages of using root mean squared error?

It punishes errors and it returns units that make sense.

## Unsupervised Learning

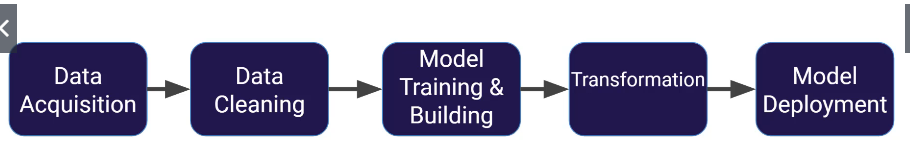
### What is unsupervised learning?

This is unlabelled data. For example, we just have images that aren’t labelled dog or cat.

### What are types of unsupervised learning?

* Clustering; cluster data into similar categories that the model deems important
* Anomaly detection; fraudulent transactions
* Dimensionality reduction; data processing techniques to reduce the number of features in a data set

### What is work flow of unsupervised learning?



No need for train/test split.