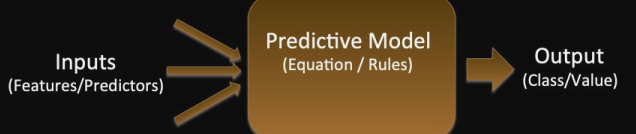
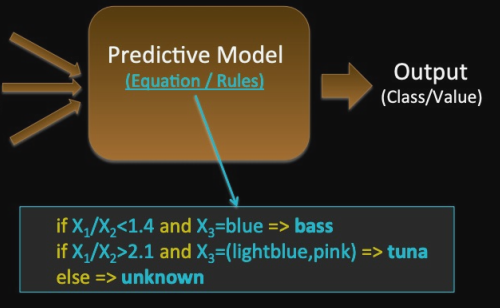
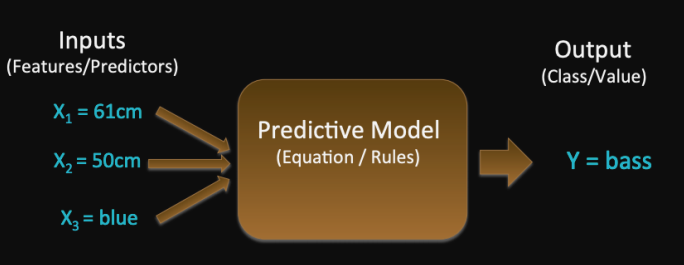
## Predictive Models -is any model that makes a prediction



* based on a set of features
* could be:
  + binary outcome (spam, not-spam
  + Categorical (bass, tuna, other)
  + real value (the age of the fish)
  + A vector of real values (probability of bass, tuna)
* If the predicted value is binary/categorical we usually refer to the model as a classifier
* If it predicts real values we refer to it as regression
* Although there are many other types of models (e.g.ranking, translation

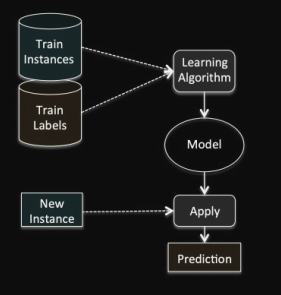
the predictive model uses equations/rules to map





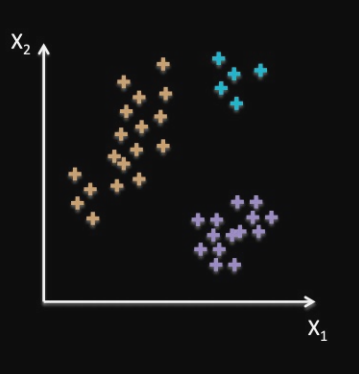
## Training a Model

Predictive models are learnt from training data and then applied to make predictions on new instances

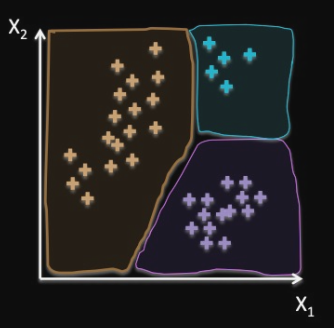


## How are models learnt?

Each instance is just a point in some feature space

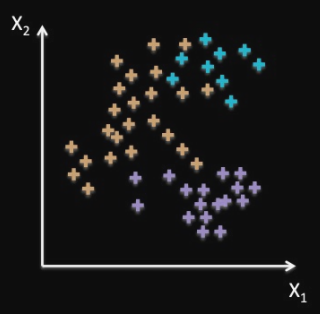


Many (classification) learning algorithms work by dividing the feature space into regions of the same type

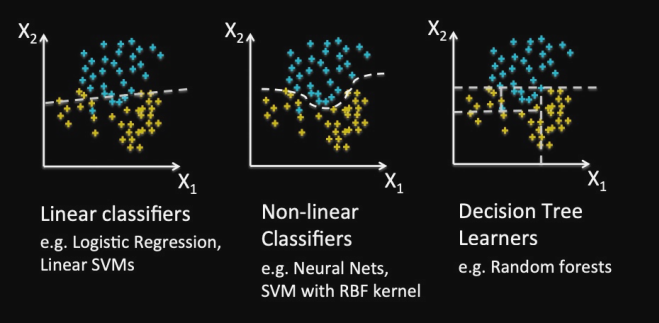


## In Practice

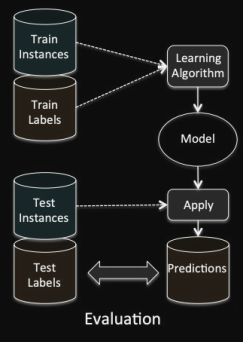
the data is usually overlapping, Making it hard to separate



Different Models

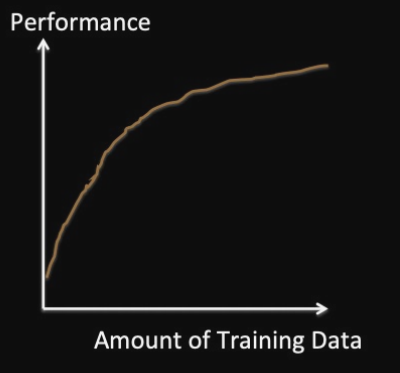


## Testing models



## Performance of models

* The more training data the better the test performance
* And (providing there is sufficient training data) the more features the better performance



# Introduction to Machine Learning

Computers to learn and act like humans do, improve by feeding data

Develop model?

►Choose a measure of success

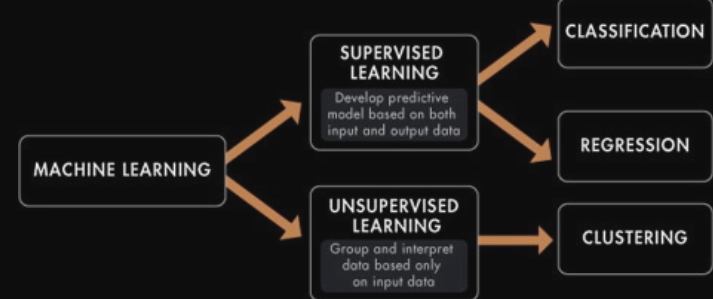
►Setting an evaluation protocol

►Developing a Benchmark Model

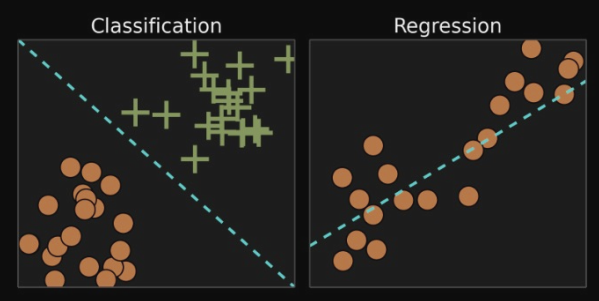
►Developing a Better Model and tunning its Hyperparameters

## Supervised vs Unsupervised Machin Learning

### Learning Styles



### Supervised



* All data is labelled and the algorithms learn to predict the output
* The goal is input data (x) we can get predict output Y
  + Classification output is a category
  + Regression output real value e.g., dollars
    - Linear regression for regression
    - Random forest for classification and regression
    - Support vector machines for classification

### Unsupervised

* All data is unlabelled and the algorithms learn to inherent structure from the input data
* The goal is finding underlying structure or distribution in the data to learn more about the data
  + Clustering finds inherent groupings in the data
  + Association Discover rules that describe large portions of your data (e.g. people that buy X also tend to buy Y)
    - k-means for clustering
    - Apriori algorithm for association

## Theory of Data Analysis

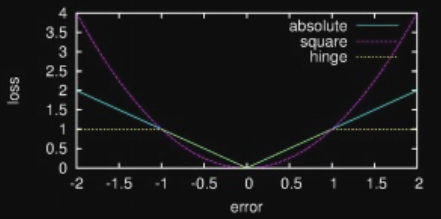
### Quality

* May be the quality of your prediction
* May be the consequence of your actions
* Can be measured on a positive or negative scale

Loss: positive when things are bad, negative (or zero) when they’re good

Gain: positive when things are good, negative when they’re not

Error: measure of “miss”, sometimes a distance, but not a measure of quality

* 
* the distance between the prediction and the actual value
* 0 means no error, prediction was exactly right

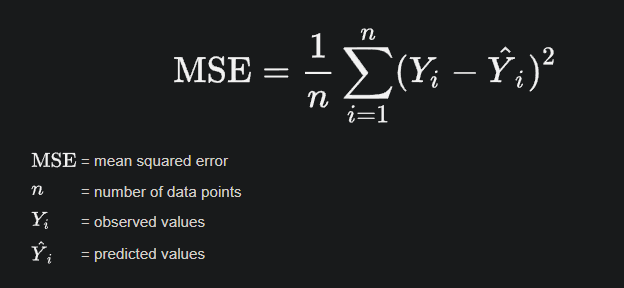
### Regression

#### Linear Regression

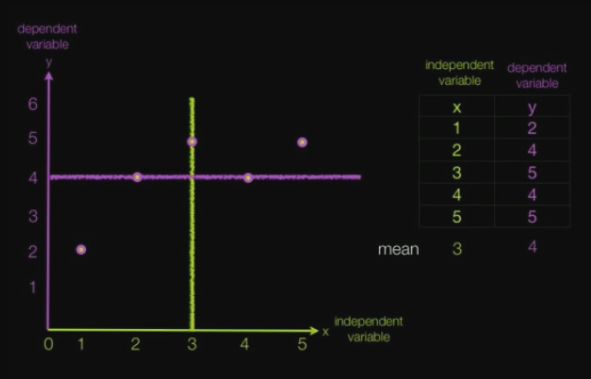
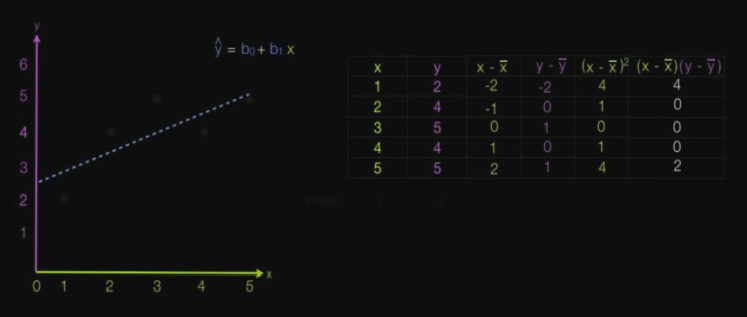
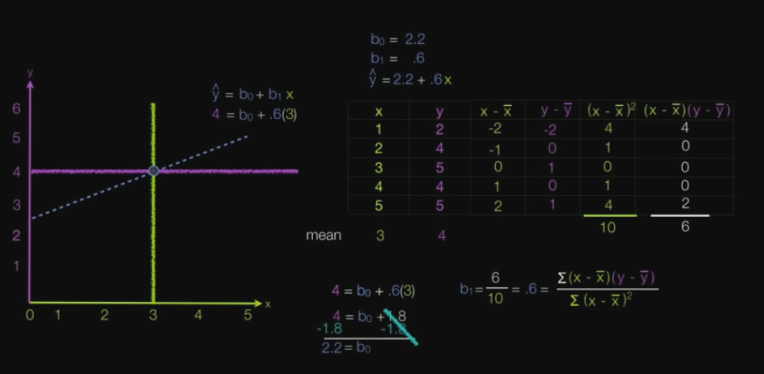
Regression fits a very simple equation to the data 

for prediction for y at the point x using the model 

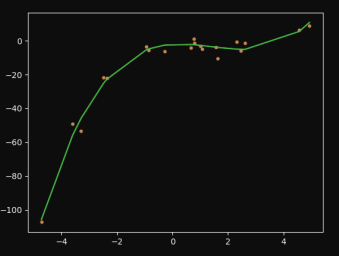
LOSS model mean square error



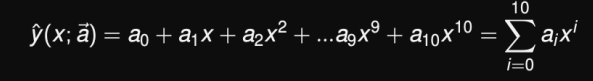
Step:

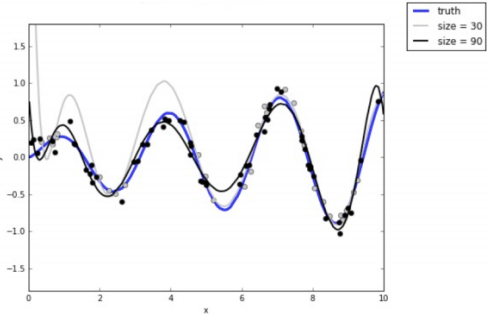
Polynomial Regression



Polynomial regression uses the same linear regression infrastructure to fit a higher order polynomial

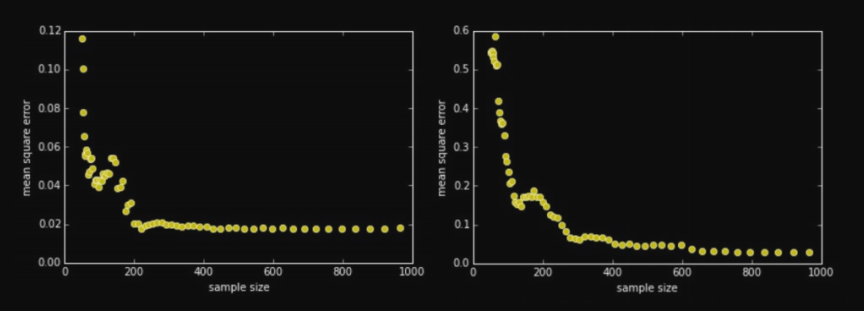
 

#### More Data Improves the Fit



* Blue line is true model that generated the data
* Grey curve is model fit to 30 data points
* Black curve is model fit to 90 data points

#### Loss decreases with Training Data



* MSE decreases as the amount of training data grows
* Different learning algorithms exhibit different behaviour