MODULE 05

DATA SCIENCE BOOTCAMP

Introduction to Machine Learning & Feature Scaling

Vertical Institute

Reminder

1. Please turn on Zoom camera the whole duration of classes.

2. At the start of all classes, please rename yourselves to: Name + Last 3 digits and letter of your NRIC. Example: John Tan (123A)

Agenda

- Introduction to Machine Learning
- Applications of Machine Learning
- Types of Machine Learning
- Bias-Variance Trade-Off
- Linear Regression
- Feature Scaling

Attendance Photo Taking

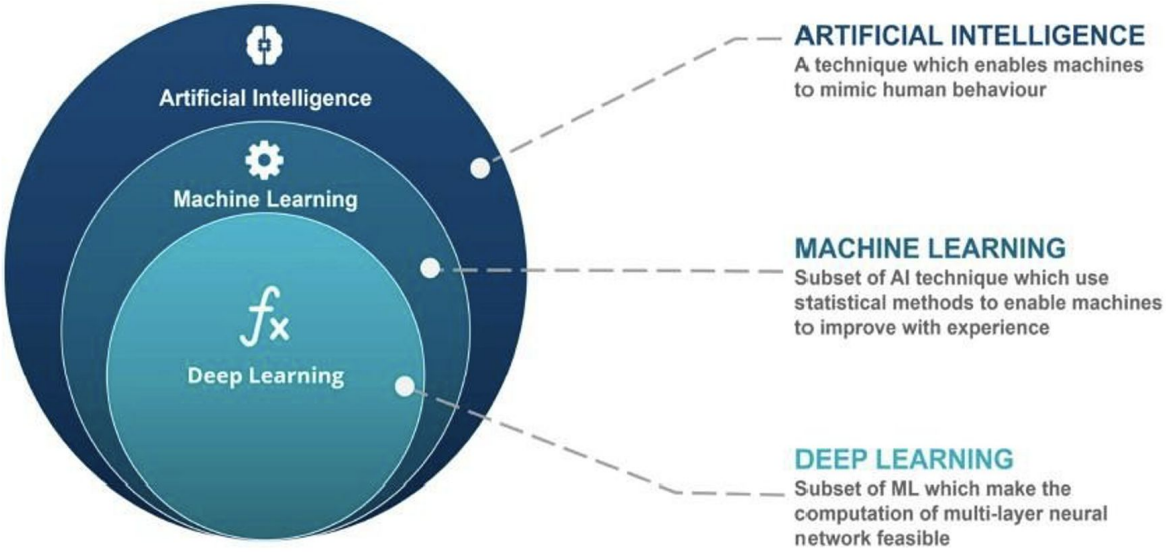
MODULE 5: INTRODUCTION TO MACHINE LEARNING & FEATURE SCALING

Machine Learning

Machine Learning Brief History

Decade	Summary
1950s	Alan Turing created “Turing Test” to determine if a computer has real intelligence. Arthur Samuel wrote the first computer learning program, the game of checkers. IBM further improved on it. Frank Robsenblatt designed the first neural network for computers
1960s	The “nearest neighbor” algorithm was written, allowing computers to begin using very basic pattern recognition
1970s	Students at Stanford University invent the “Stanford Cart” which can navigate obstacles in a room on its own.
1980s	Gerald Dejong introduces the concept of Explanation Based Learning (EBL), in which a computer analyses training data and creates a general rule it can follow by discarding unimportant data. Terry Sejnowski invents NetTalk, which learns to pronounce words the same way a baby does
1990s	Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions. IBM’s Deep Blue beats world champion at chess.
2000s	Geoffrey Hinton coins the term “deep learning” to explain new algorithms that let computers “see” and distinguish objects and text in images and videos
2010s	Deep learning becomes feasible, which leads to machine learning become integral to many widely used software services and applications.

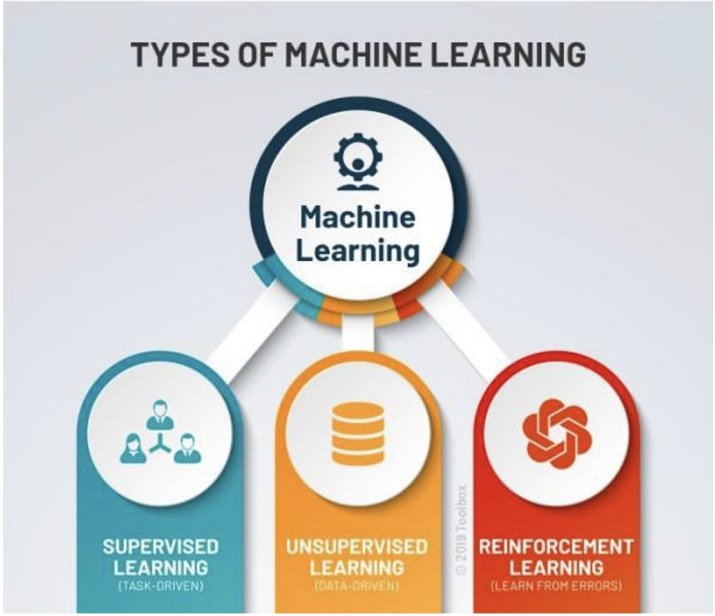
AL – ML – DL in Short



Applications For ML

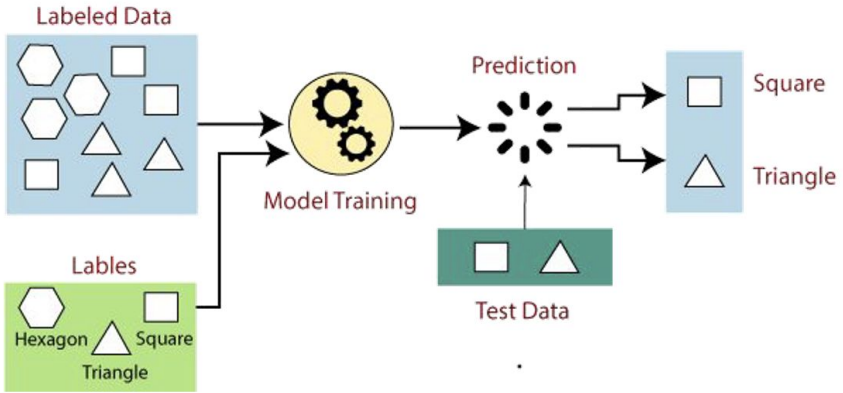
- Image Recognition – Face detection
- Speech recognition – Google Assistant, Siri, Cortana, Alexa
- Traffic Prediction – Real time Location
- Product Recommendations – Amazon, Netflix
- Email Spam and Malware filtering
- Online Fraud Detection

Types Of Machine Learning



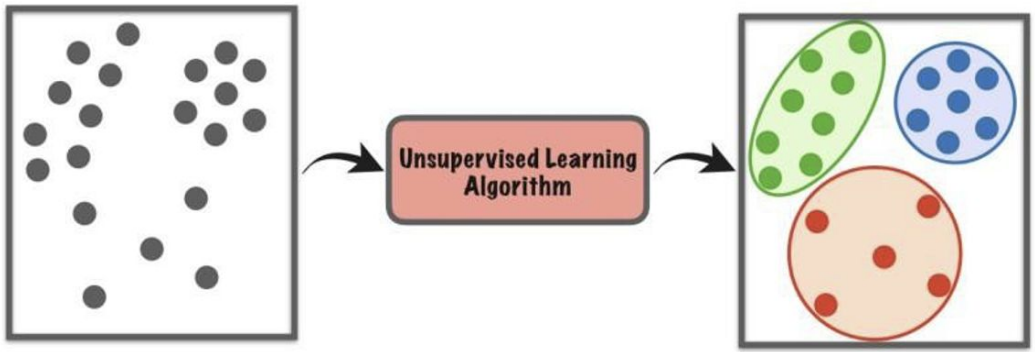
Supervised Learning

- “Supervised” refers to the fact that labeled training data is required.
- During training, the model creates an internal representation that represents a generalization of the training data, and uses it to make predictions of the labels of new data.



Unsupervised Learning

- Unsupervised learning methods are used when there is no labeled data. In other words, there is no ground truth.
- Instead of predicting a class as we do with supervised learning, data points are grouped based on measures of similarity.

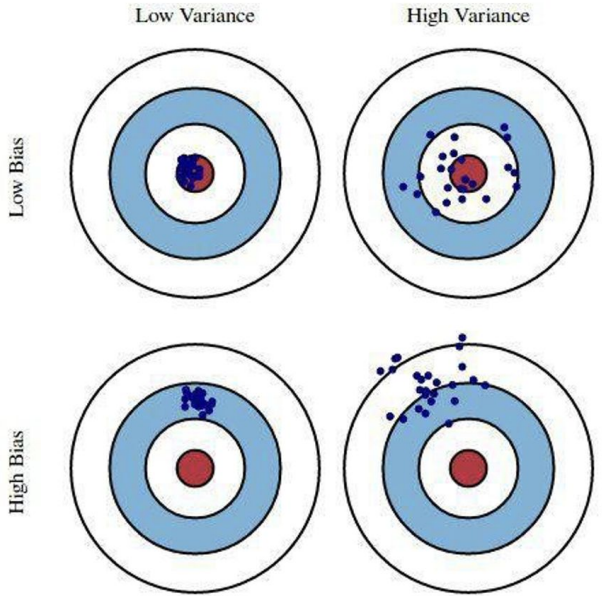


Reinforcement Learning

- Reinforcement learning is a reward-based learning technique.
- Different from supervised learning since no correct input/output pairs need to be present for training
- Is considered a separate learning paradigm from supervised and unsupervised learning.

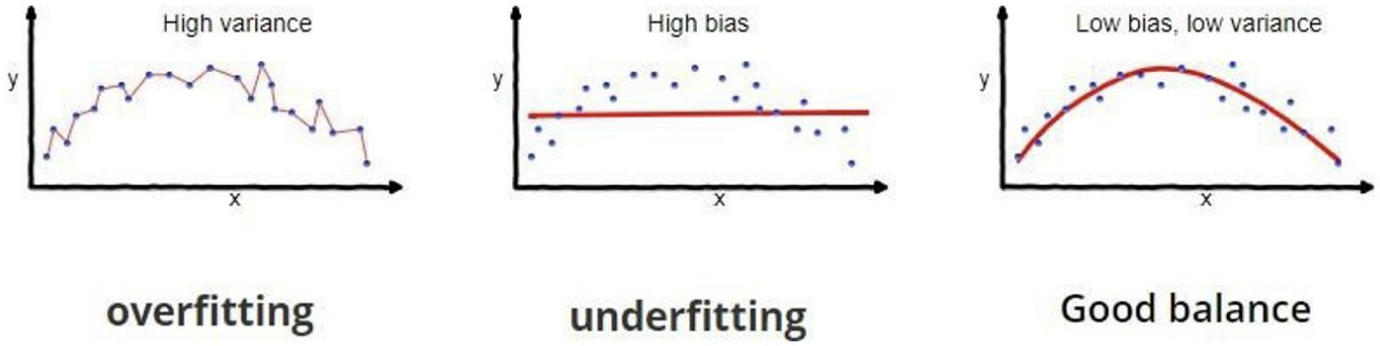
Bias-Variance Trade-Off

- Bias is the difference between the average prediction of our model and correct value is hard to predict
- Variance is the variability of model prediction for a given data point or a value which tells the spread of our data



Bias-Variance Trade-Off

- **Underfitting** happens when a model is unable to capture the underlying pattern of the data
- **Overfitting** happens when our model captures the noise along with the underlying pattern in the data.



Universal Machine Learning Workflow

1. Define the problem and assemble a dataset
2. Choose a measure of success
3. Decide on an evaluation protocol
4. Prepare the data
5. Develop a model that does better than a baseline
6. Develop a model that overfits
7. Regularize the model and tune its hyperparameters

Actual AI Practice in Financial Institutions

"Using a machine learning model, we are able to then rate the probability of the next transaction being fraudulent"

- Mr Royston Soon, VP of fraud risk management at OCBC Bank.

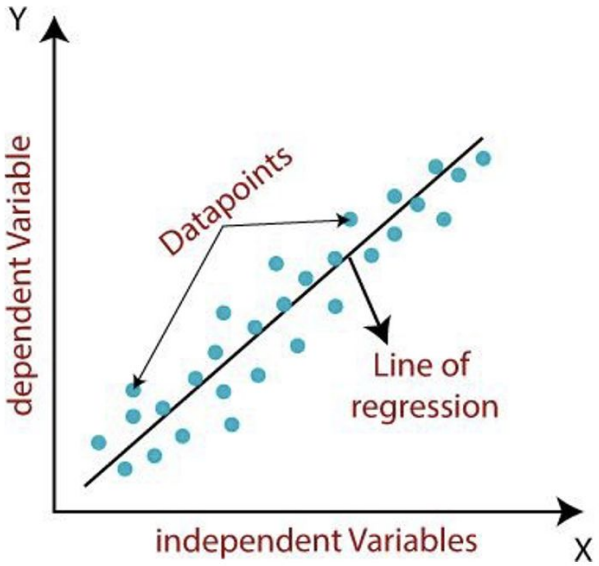
https://www.straitstimes.com/tech/tech-news/anti-fraud-experts-use-ai-to-predict-cheaters-next-move

MODULE 5: INTRODUCTION TO MACHINE LEARNING & FEATURE SCALING

Supervised Learning: Linear Regression

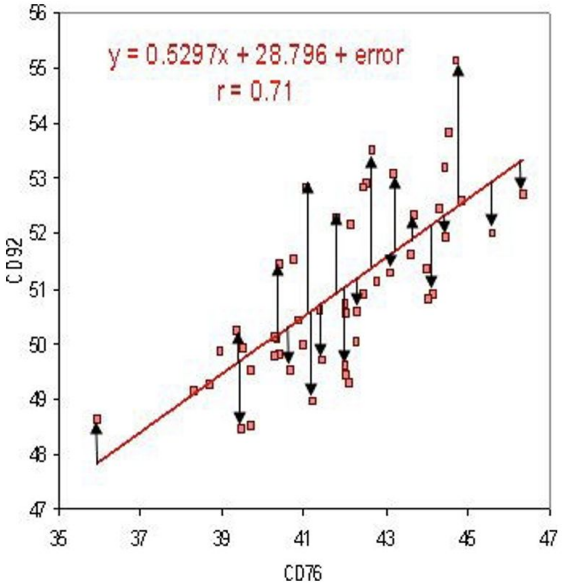
Supervised Learning - Regression

- Regression analysis is a set of statistical methods applied to estimate the relationship among variables.
- Regression is used when the target outputs are continuous, i.e. numerical outputs are expected
- It is used to determine if a correlation exists between variables



Evaluation of Regression Model

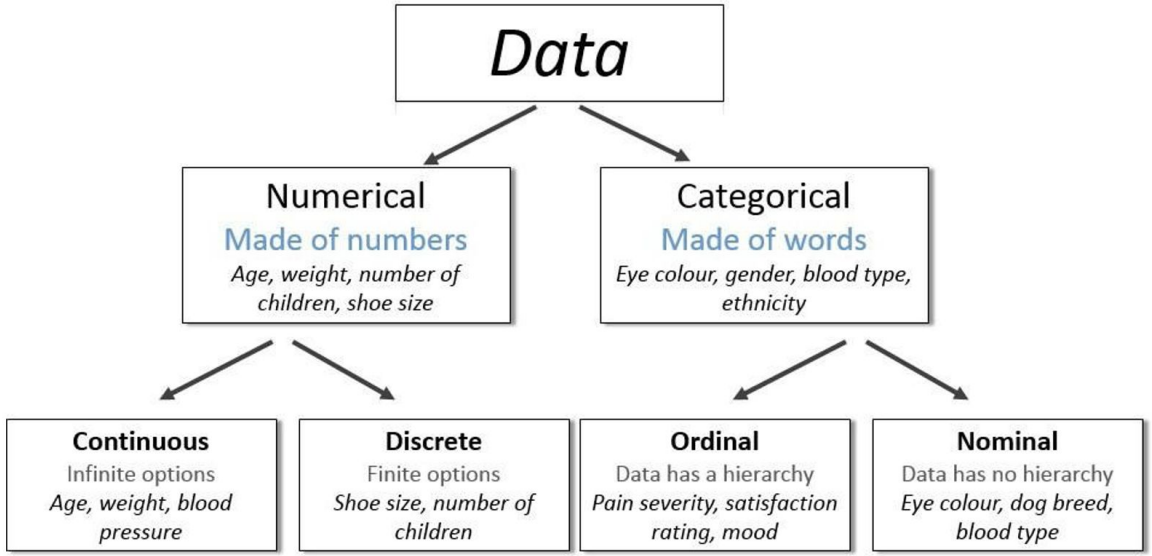
- Mean Squared Error (MSE) = Σ(yi – y' i)2 / n
- Root Mean Squared Error (RMSE) = √MSE
- Mean Absolute Error (MAE) = Σ |yi – y' i| / n



MODULE 5: INTRODUCTION TO MACHINE LEARNING & FEATURE SCALING

Working With Categorical Data

Remember Types of Data?



Problems With Categorical Data

- Some algorithms deal with categorical data directly while some cannot deal with them directly (meaning they have to be numeric)

- In general, due to the efficient implementation of the various algorithms, it is required that categorical data be converted to a numeric form.

Integer Encoding

- Used for ordinal relationship
- A column named Class with values ‘first’, ‘second’, and ‘third’

Class_column	Encoded_column
First	1
Second	2
Third	3

One-Hot Encoding

- Used when no ordinal relationship exists between values

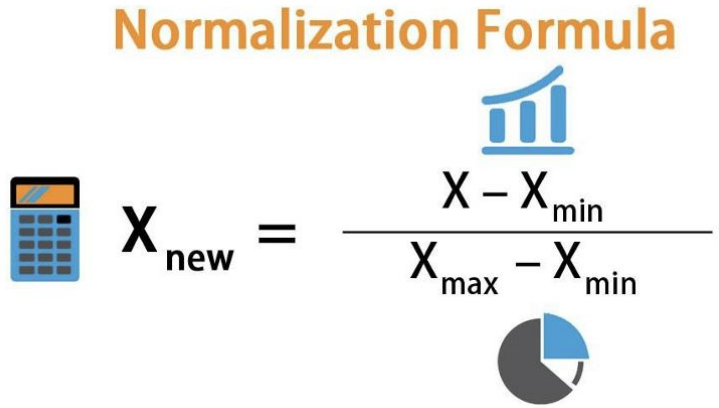
Red	Green	Blue
1	0	0
0	1	0
0	0	1

So...

- Red will be represented as: [1 0 0]
- Green will be represented as: [0 1 0]
- Blue will be represented as: [0 0 1]

Feature Scaling

- Feature scaling is a method used to normalize the range of independent variables or features of data.
- Also known as data normalization is generally performed during the data pre-processing step
- Normalization (also known as Min-Max scaling) rescales the values into a range of [0,1]



Uses of Normalization

- The goal is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.
- Consider a dataset containing 2 features, age and income. Age ranges from 1 to 100
- Income ranges from 20,000 to 500,000
- The attributed income will intrinsically influence the result more due to its larger value but it does not mean it is a more important predictor.

Attendance Photo Taking

MODULE 05

DATA SCIENCE BOOTCAMP

Thank you!

Vertical Institute