18CSE392T – Machine Learning I

Department of Data Science and Business Systems

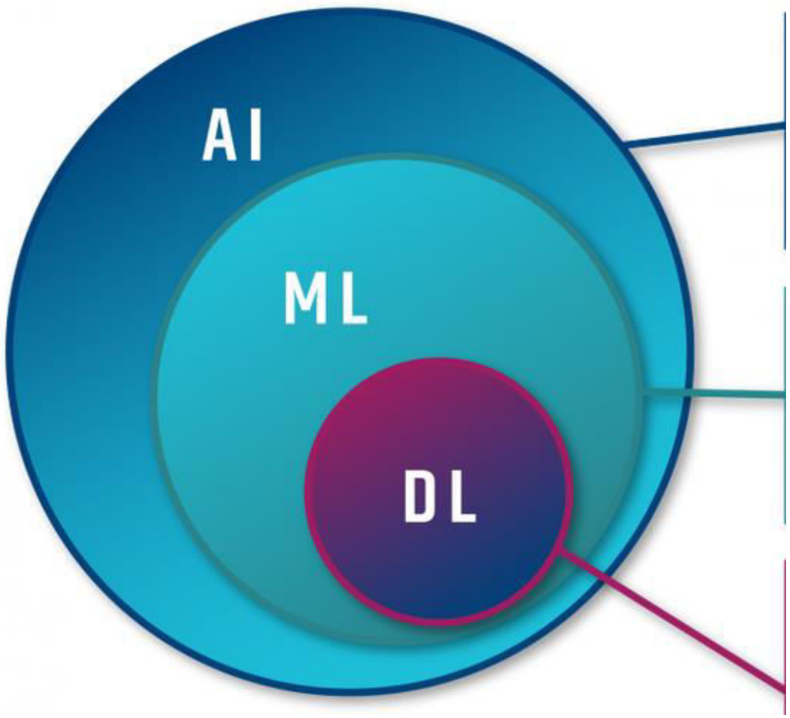
Agenda

1 *What is ML ?*

2 *Why do we need ML?*

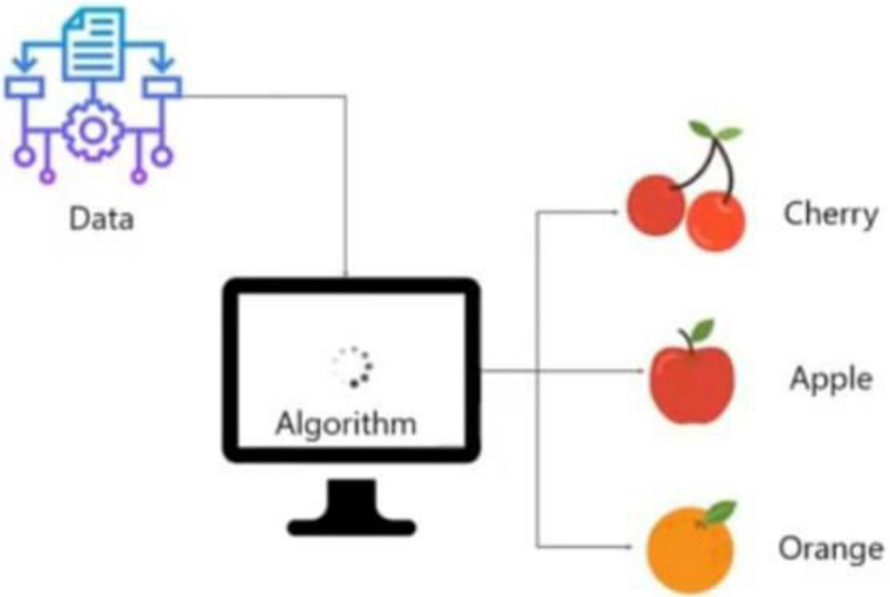
3 *Where do we use it ?*

Introduction



- *Artificial Intelligence*
- *Machine Learning*
- *Deep Learning*

Child Learning



Learning

- Past experience
- Learning from others

- Machines need instruction from humans

Machine Learning – In general

- A subset of Artificial Intelligence
- Machines that learn automatically & improve from experience.
- No explicit programme is needed.

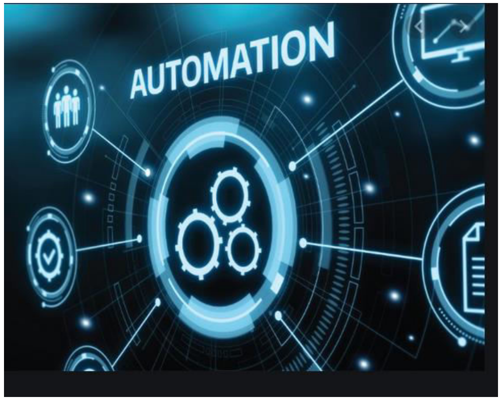
Definition - ML

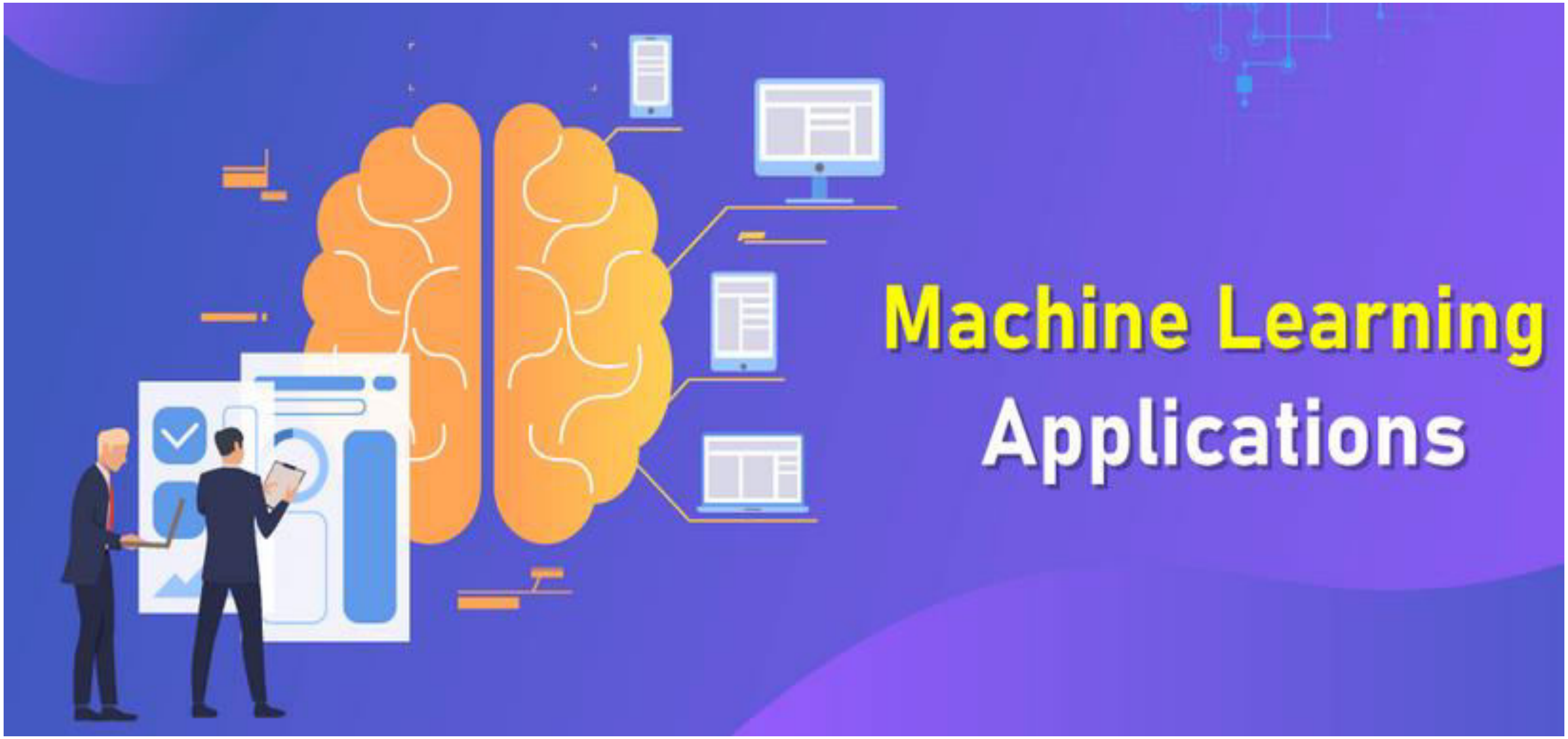
- A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.

Need for ML

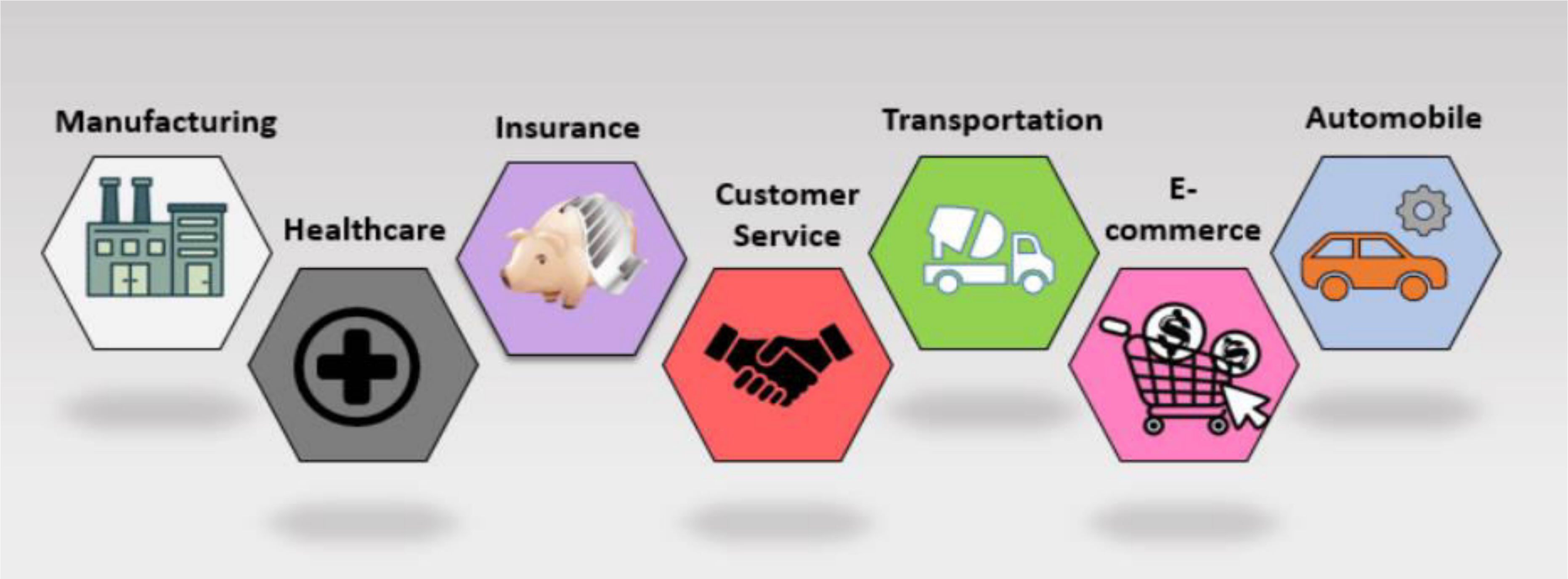
- Huge amount of data (Big data)

- Automated methods of data analysis




Machine Learning
Applications

Few ML Applications




Thank
you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

Machine Learning Types

- Supervised
- Unsupervised
- Reinforcement

SUPERVISED LEARNING

Agenda

- Definition
- Data
- Categories
- Supervision of training
- Applications
- Algorithms

Supervised Learning

1. Labelled Data
2. Direct Feedback
3. Predict outcome / future

Definition

- Supervised ML is a method in which we teach the machines using labelled data.

In other words.....

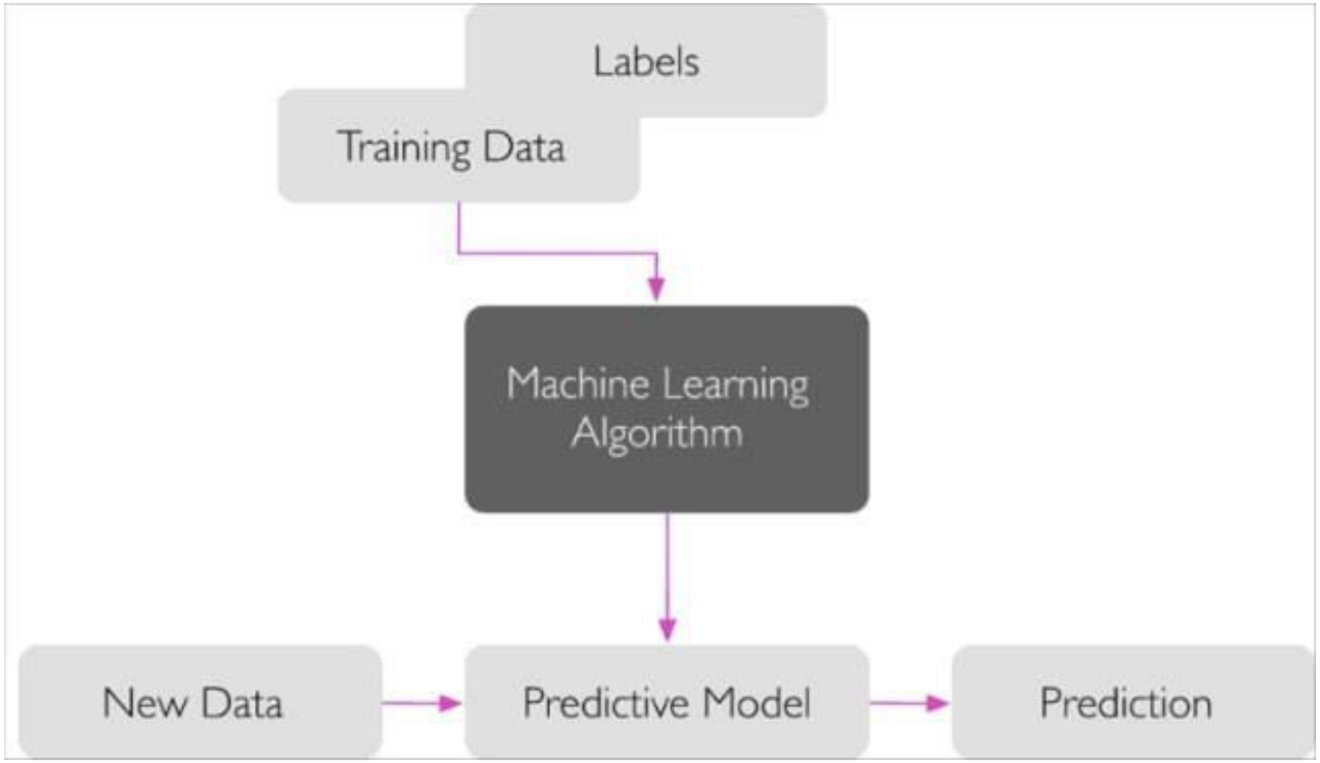
- The main goal in supervised learning is to learn a model from labeled training data that allows us to make predictions about unseen or future data.

Data

- Labelled Input Data
- Labelled Output Data

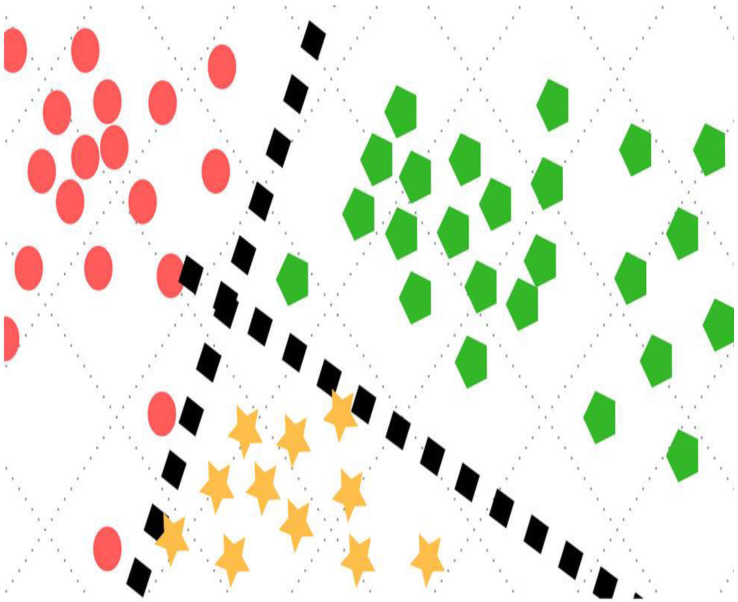
Supervision of Training

- Needs External Supervision from training data

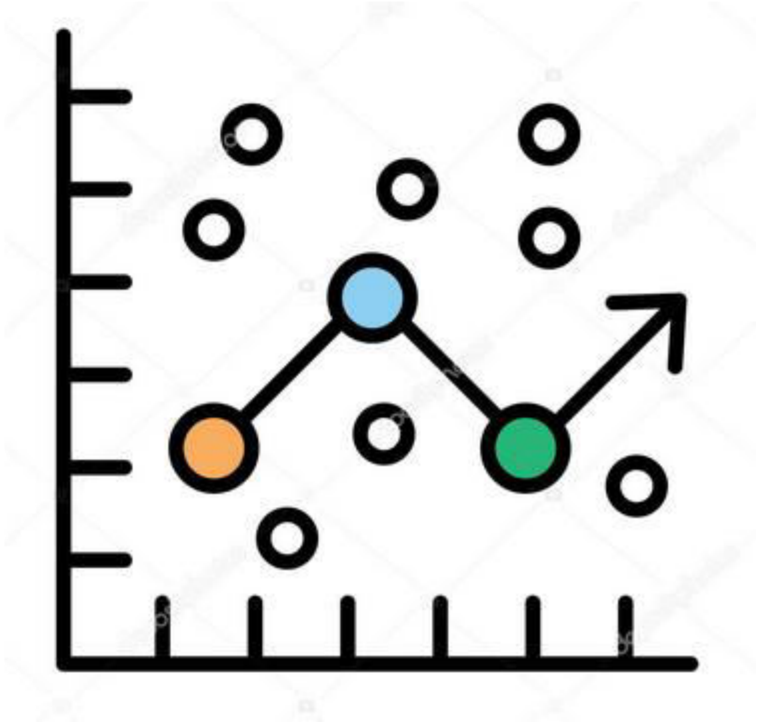


Categories

Classification
Input : Discrete



Regression
Input : continuous



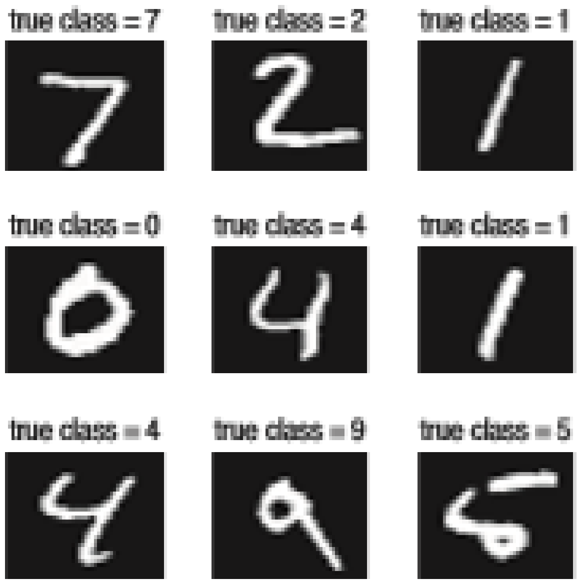
APPLICATIONS (SUPERVISED LEARNING)

Email Spam Filtering

- A corpus of labeled emails, emails that are correctly marked as spam or not-spam
- To predict whether a new email belongs to either of the two categories

Handwriting Character recognition

- **Multiclass classification** task



Predicting Real Estate Price

Weather Forecasting

- Weather forecasting is a prediction of what the weather will be like in an hour, tomorrow, or next week.

Algorithms

- Linear Regression
- Logistic Regression
- K- Nearest Neighbor
- Bayesian Classification
- Decision Trees
- Support Vector Machine (SVM)
- Random Forest


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

UNSUPERVISED LEARNING

Unsupervised Learning

1. Un labelled data
2. No guidance
3. Relatively co-occur

Definition

- Unsupervised ML is a method in which the machine is trained on unlabelled data without any guidance.

In other words.....

- For just given output data, without any inputs. The goal is to discover “interesting structure” in the data; this is sometimes called **knowledge discovery**.

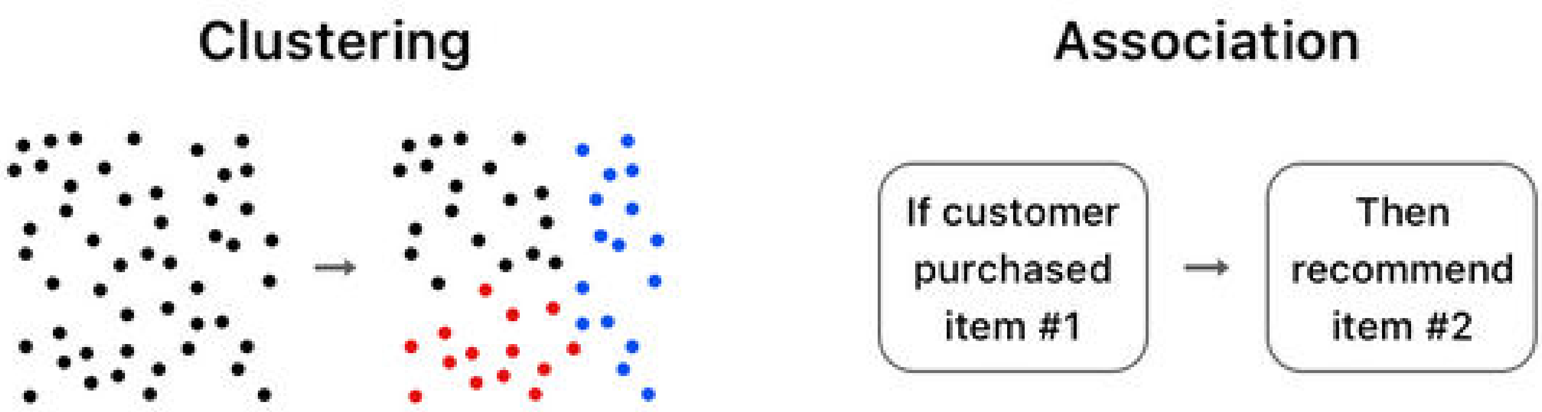
Categories

Clustering

- Grouping based on similarity

Association

- Discovering patterns in data



Association

Association rules allow you to establish associations amongst data objects inside large databases.

This unsupervised technique is about discovering interesting relationships between variables in large databases.

For example, people that buy a new home most likely to buy new furniture.

Clustering

Clustering: is the assignment of a set of observations into subsets (called **clusters**) so that observations in the same **cluster** are similar in some sense.

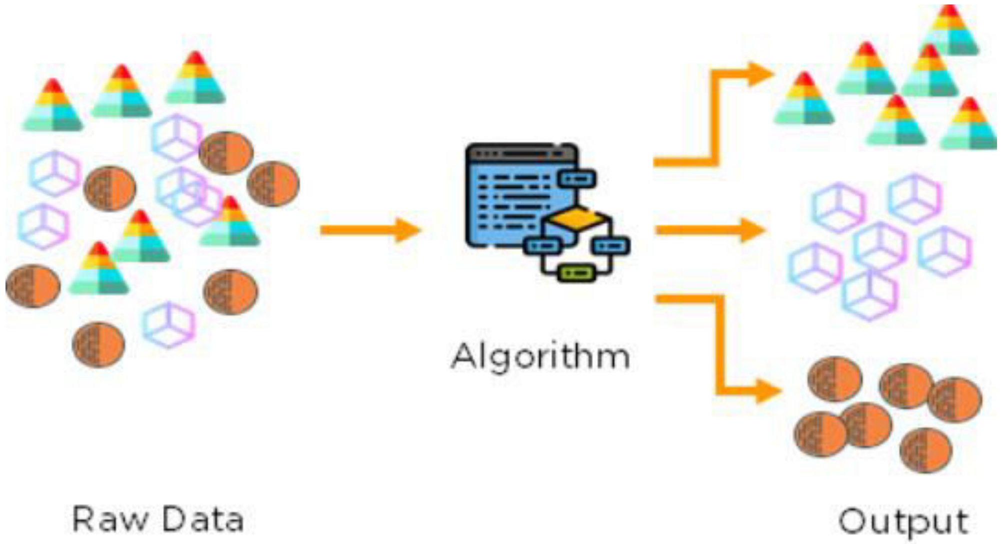
Clustering is a method of unsupervised **learning**, and a common technique for statistical data analysis used in many fields.

Data

- Unlabelled Input Data
- Patten Output Data (grouping)

Unsupervision of Training

- Needs **NO External Supervision** from training data



APPLICATIONS (SUPERVISED LEARNING)

Vegetable clustering

Clustering

sample

Cluster/group

Identifying fraudulent or criminal activity

- In this scenario, we are going to focus on fraudulent taxi driver behavior.

- **What is the problem:** You need to look into fraudulent driving activity. The challenge is how do you identify what is true and which is false?

- **How clustering works:** By analysing the GPS logs, the algorithm is able to group similar behaviors. Based on the characteristics of the groups you are then able to classify them into those that are real and which are fraudulent.

Algorithms

- K – means
- C- means
- Apriori algorithm


Thank you

18CSE392T – Machine Learning I

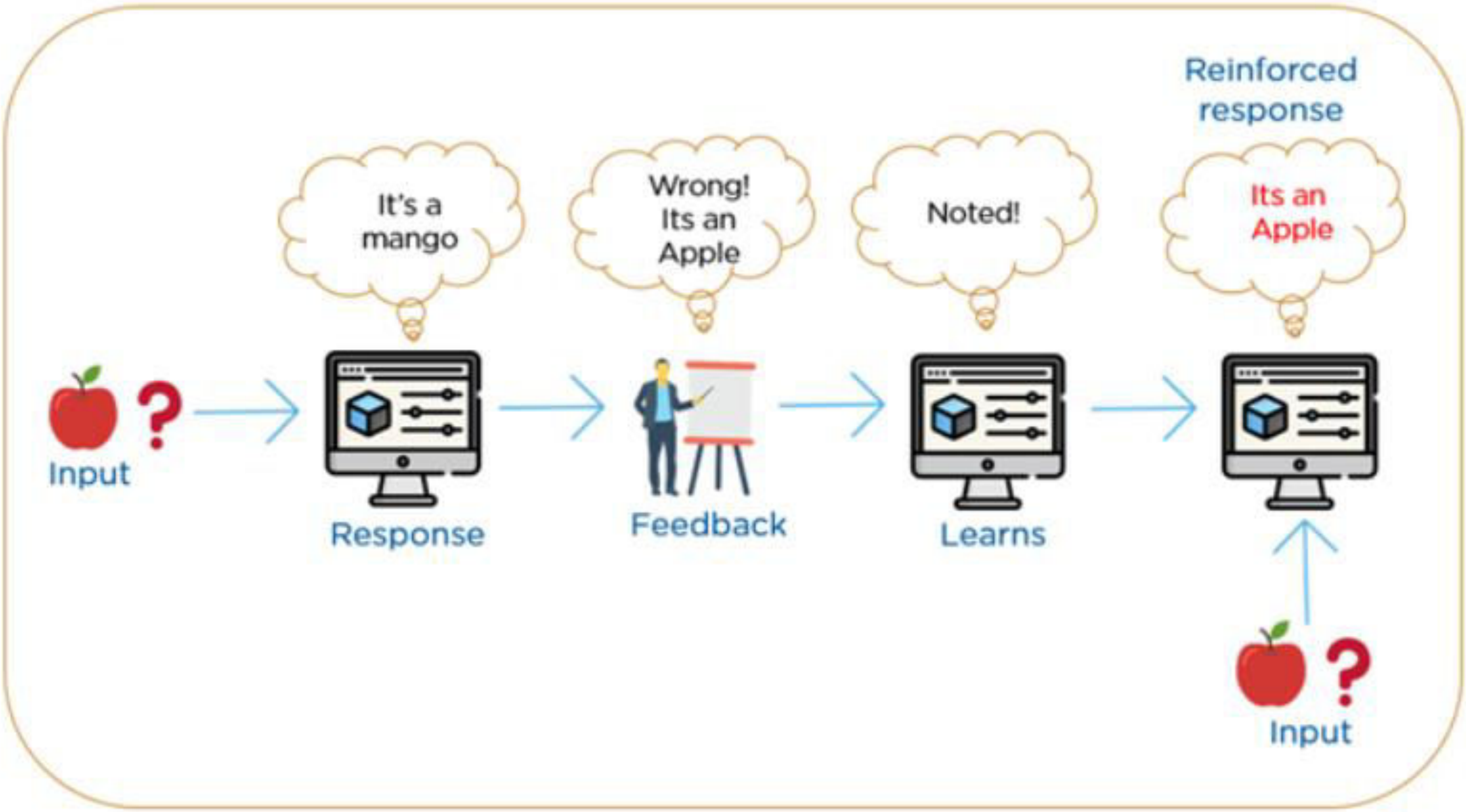
Department of Data Science and Business Systems

Reinforcement Learning

Definition

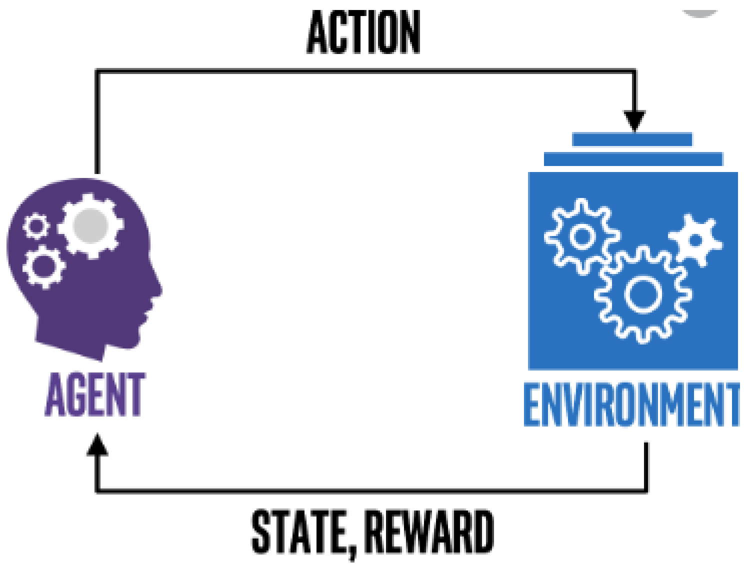
- Reinforcement learning (RL) is a type of ML which is all about taking suitable action to maximize reward in a particular situation.
- It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.
- RL means to establish or encourage a pattern of behavior.

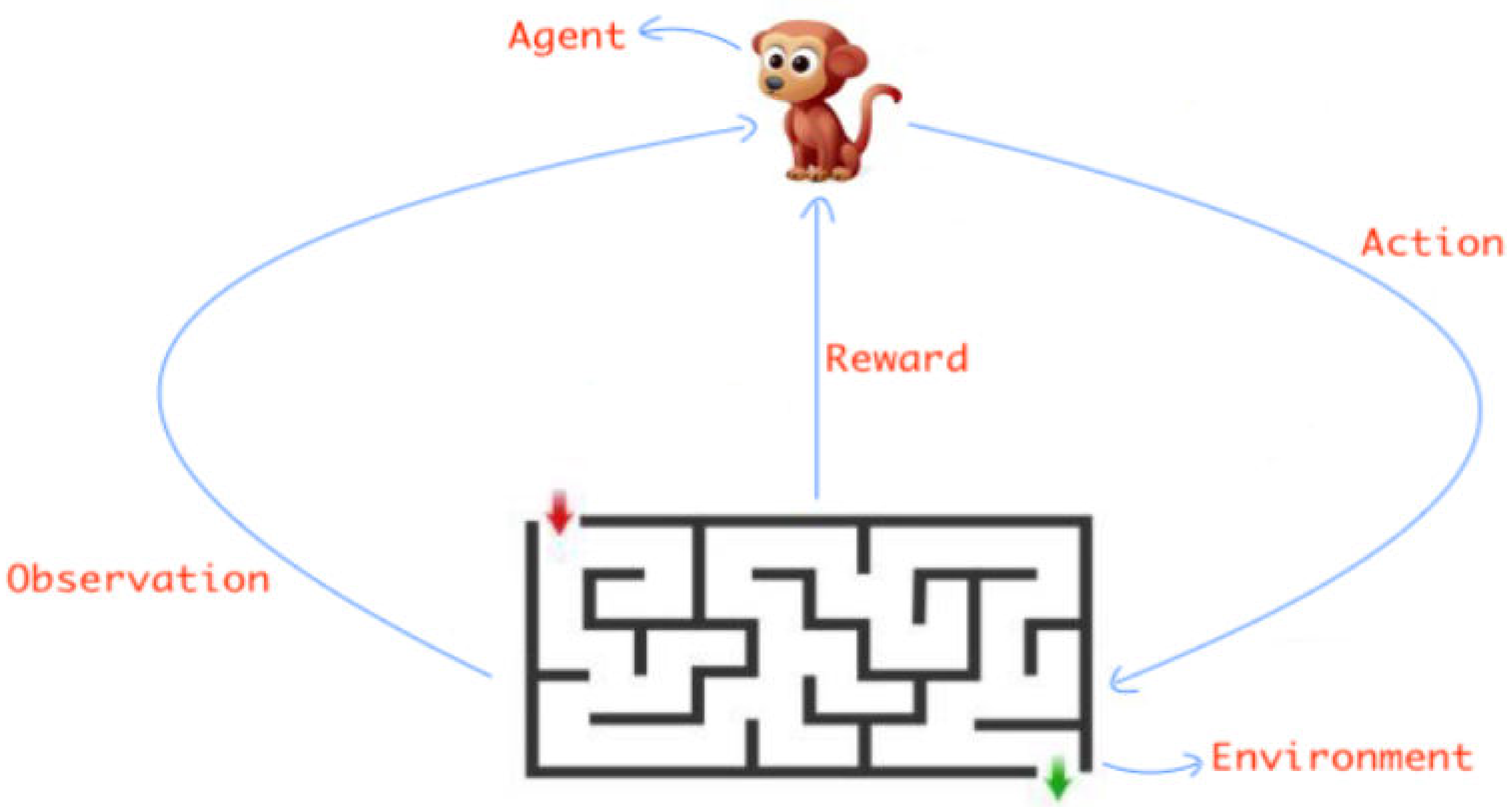
Reinforcement Learning is a subfield of machine learning that teaches an agent how to choose an action from its action space, within a particular environment, in order to maximize rewards over time.


Reinforced response
It's a mango
Wrong! Its an Apple
Noted!
Its an Apple
Input
Response
Feedback
Learns
Input

Essential Elements

- **Agent.** The program you train, with the aim of doing a job you specify.
- **Environment.** The world, real or virtual, in which the agent performs actions.
- **Action.** A move made by the agent, which causes a status change in the environment.
- **Rewards.** The evaluation of an action, which can be positive or negative.


ACTION
AGENT
ENVIRONMENT
STATE, REWARD


Agent
Action
Reward
Observation
Environment

RL Works

- Observation of the environment
- Deciding how to act using some strategy
- Acting accordingly
- Receiving a reward or penalty
- Learning from the experiences and refining our strategy
- Iterate until an optimal strategy is found

Placement of Ads

- **Agent:** The program making decisions on how many ads are appropriate for a page.
- **Environment:** The web page.
- **Action:** One of three: 1. Putting another ad on the page. 2. Dropping an ad from the page. 3. Neither adding nor removing.
- **Reward:** Positive when revenue increases; negative when revenue drops.

Controlling A Walking Robot

- **Agent:** The program controlling a walking robot.
- **Environment:** The real world.
- **Action:** One out of four moves

1. Forward 2. Backward 3. Left and 4. Right.

- **Reward:** Positive when it approaches the target destination; negative when it wastes time, goes in the wrong direction or falls down.
- In this final example, a robot can teach itself to move more effectively by adapting its policy based on the rewards it receives.


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

Curse of Dimensionality

Curse of Dimensionality

Dimension

Features

Attributes

SCENARIO

Real estate rate prediction

Features

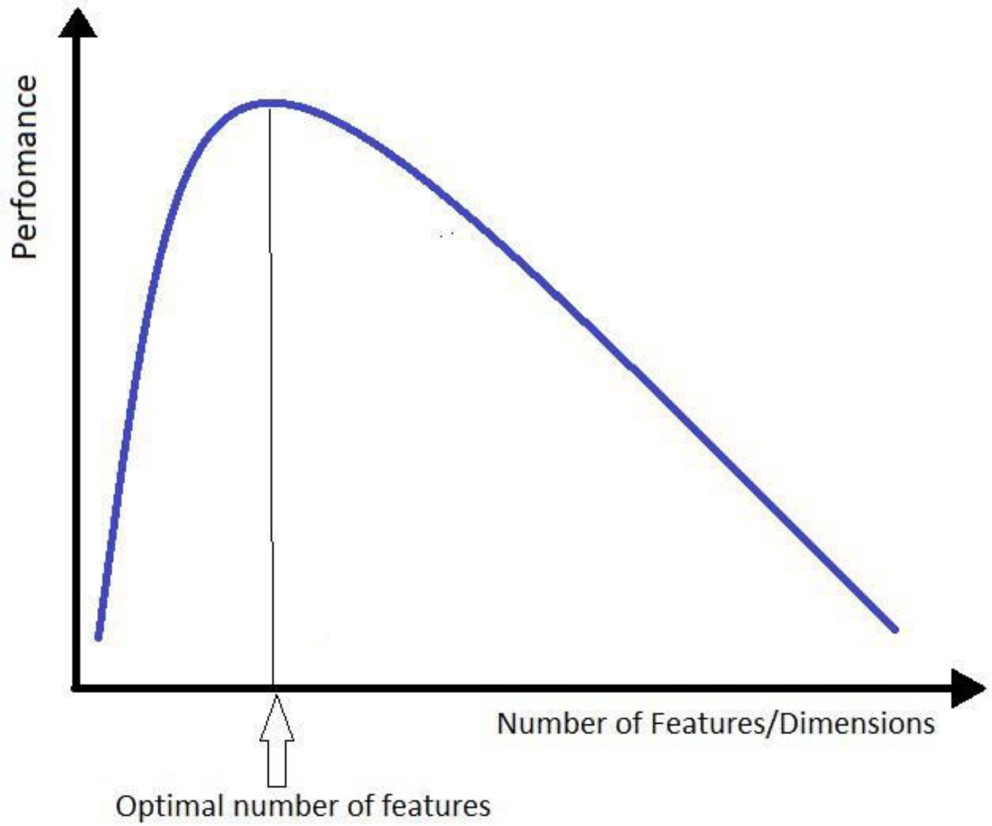
- House age
- No. of rooms
- No. of bed rooms
- No. of floors
- Type
- Location
- Cost

Analyze the performance of a Formula One (F1) driver

i) **Model_1** consists of only two features say the circuit name and the country name.
ii) **Model_2** consists of 4 features say weather and max speed of the car including the above two.
iii) **Model_3** consists of 8 features say driver's experience, number of wins, car condition, and driver's physical fitness including all the above features.
iv) **Model_4** consists of 16 features say driver's age, latitude, longitude, driver's height, hair color, car color, the car company, and driver's marital status including all the above features.
v) **Model_5** consists of 32 features.
vi) **Model_6** consists of 64 features.
vii) **Model_7** consists of 128 features.
viii) **Model_8** consists of 256 features.
ix) **Model_9** consists of 512 features.
x) **Model_10** consists of 1024 features.

Model_4 don't actually contribute anything towards analyzing the performance of the F1 driver.
For example, the driver's height, hair color, car color, car company, and the driver's marital status is giving useless information for the model to learn, hence the model gets confused with all this extra information, and the accuracy starts to go down.

Features



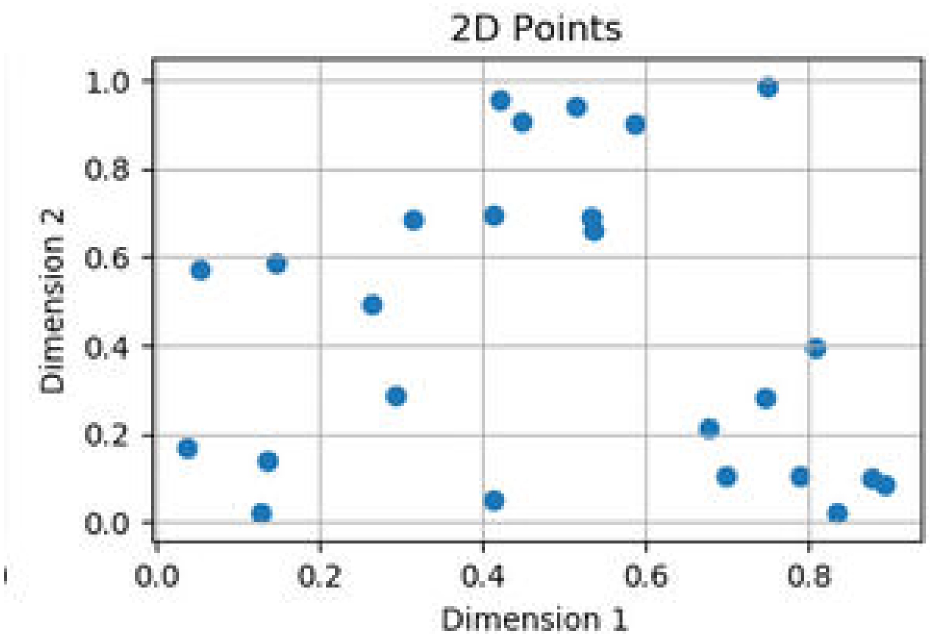
Curse of Dimensionality describes the explosive nature of increasing data dimensions and its resulting exponential increase in computational efforts required for its processing and/or analysis.

One Dimension

Number of Rooms

Two Dimension

Number of Rooms vs Costs



N- Dimension

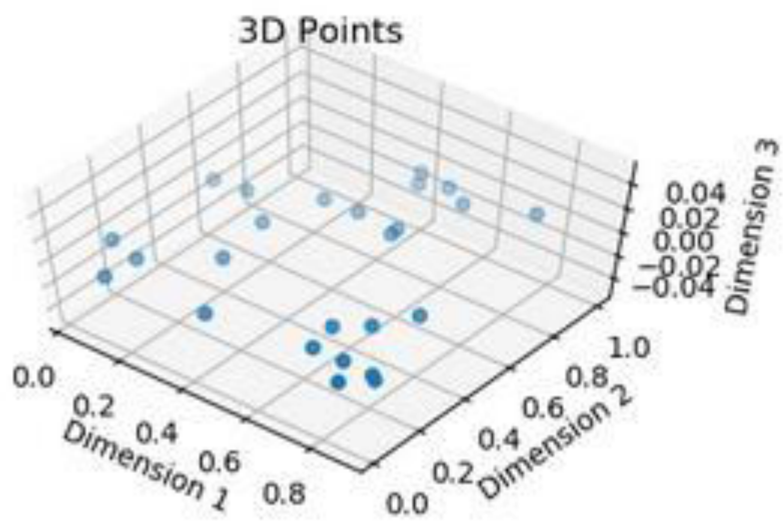
Number of Rooms

Vs

Number of Floors

Vs

Costs



Definition

The curse of dimensionality refers to the phenomena that occur when classifying, organizing, and analyzing high dimensional data that does not occur in low dimensional spaces, specifically the issue of data sparsity and "closeness" of data.

Cont..

- Sparsity of data occurs when moving to higher dimensions.
- The volume of the space represented grows so quickly that the data cannot keep up and thus becomes sparse.

Cont..

- As the data space seen above moves from one dimension to two dimensions and finally to three dimensions, the given data fills less and less of the data space.
- In order to maintain an accurate representation of the space, the data for analysis **grows exponentially**.
- **Issues** with sorting or classifying the data. In low dimensional spaces, data may seem very similar but the higher the dimension the further these data points may seem to be.
- **Infinite Features Requires Infinite Training**

Drawbacks (Summary)

- As the dimension $\uparrow$ data become more sparse.
- Hard to generalize
- Need more training data
- If the dimension $\uparrow$,every data point is equidistant from all other points.


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

OVERFITTING
and
UNDERFITTING

Terms

Signal

In predictive modelling, Signal refers to the true underlying pattern that helps the model to learn the data

Terms

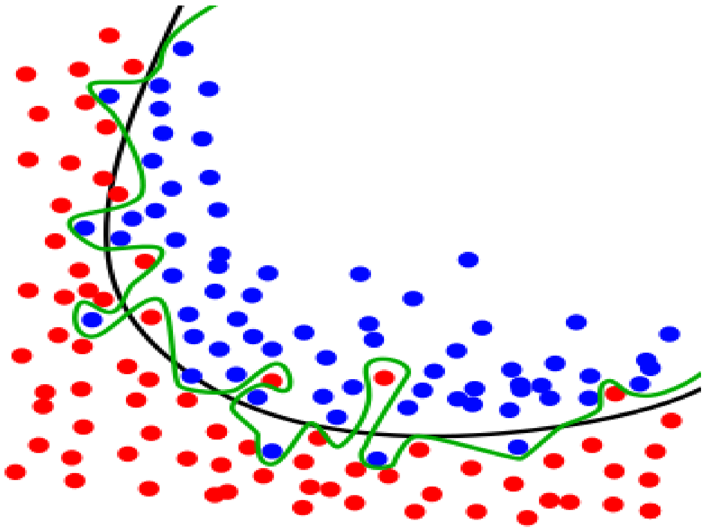
Noise

Noise is irrelevant and random data in the dataset.

Overfitting

- Overfitting refers to a model that models the training data too well.
- Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.
- This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the m odel.
- The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize.

Black line – Good fit ***Green line - Over fit***



The main challenge of overfitting is to estimate the accuracy of the performance of our model with new data

Techniques to reduce overfitting

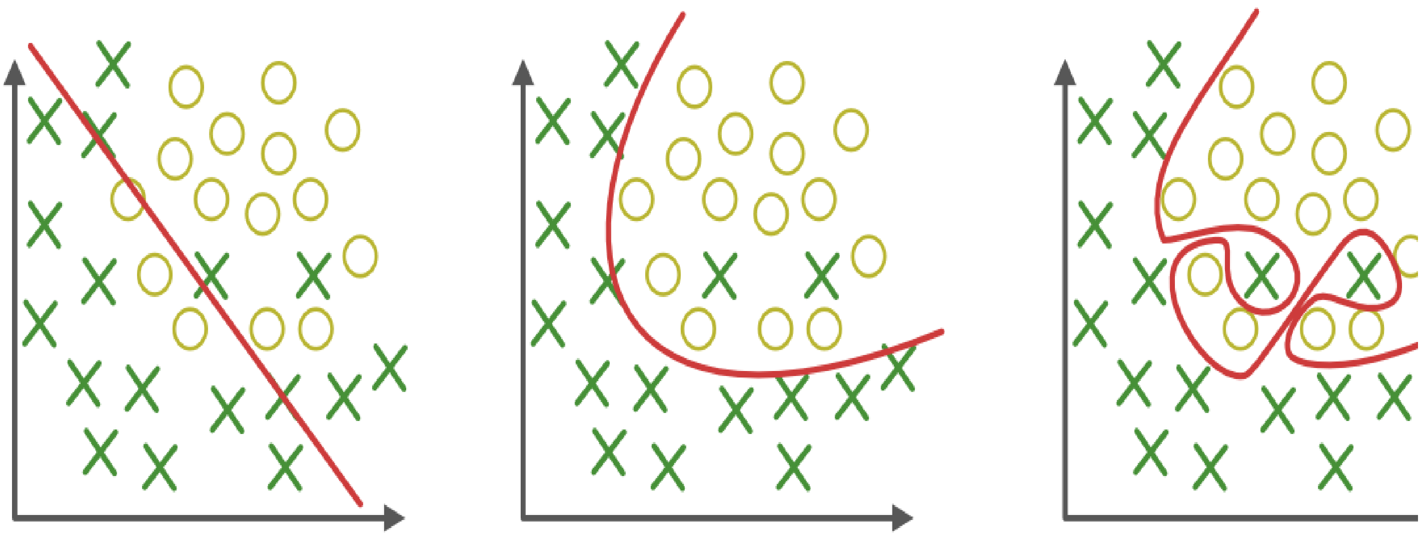
- ✓ Reduce model complexity.
- ✓ Early stopping during the training phase.
- ✓ Use dropout for neural networks to tackle overfitting

Underfitting

- Underfitting refers to a model that can neither model the training data nor generalize to new data.
- An underfit machine learning model is not a suitable model and will be obvious as it will have poor performance on the training data.

Techniques to reduce underfitting

- ✓ Increase model complexity
- ✓ Increase number of features
- ✓ Remove noise from the data.
- ✓ Increase the number of epochs or increase the duration of training to get better results.



Under-fitting

(too simple to explain the variance)

Appropirate-fitting

Over-fitting

(forcefitting--too good to be true)


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

LINEAR REGRESSION

OBJECTIVE - 1

- Linear regression analysis is used to predict the value of a variable based on the value of another variable.
- Establish if there is a relationship between two variables.

Income and Spending
Wage and Gender

OBJECTIVE - 2

- Predict for new observations.

What will be the sales in the next month?
What will be the weather the next day?

Variables

Dependent Variable	Independent variable

Dependent Vs Independent Variable

Dependent Variable

- **Variable we want to explain or forecast**
- **Depend on other variable(S)**
- **Noted by ‘y’**

Independent Variable

- **Variable that explains other variable**
- **Its values are independent**
- **Noted by ‘x’**

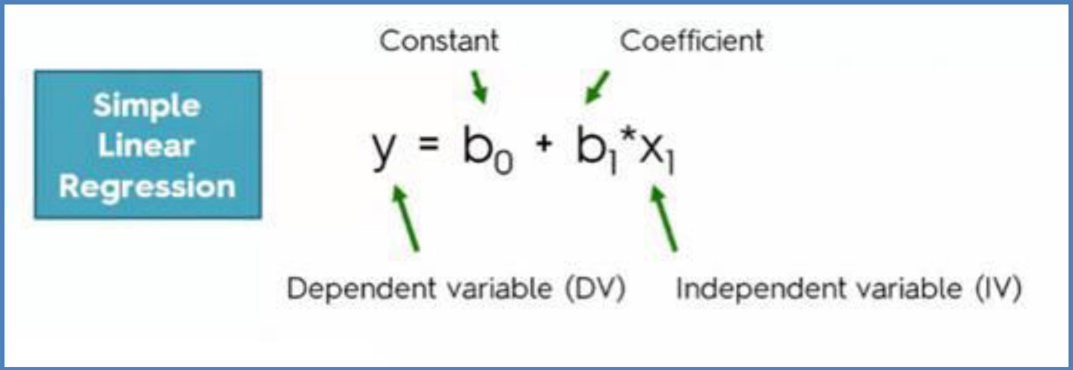
Basic Idea of Slope

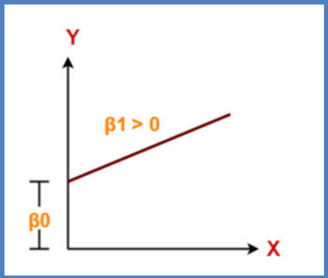
You may remember this slope formula

$$\mathbf{y = m\,x + c}$$

- m – slope
- c – y-intercept

Simple Linear Regression





Scenario - SLR

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

Linear regression estimate

Sales = 168 + 23 Advertising

Multiple Linear Regression

$$Y = b_0 + b_1x_1 + b_2x_2 + \ldots. + b_nx_n$$

Where Y = Dependent Variable(DV)

X_1, x_2, x_n – Indepdent Variable(IV)

b_0 – intercept

b_1, b_2 – coefficients

n – No. of observations


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

BIAS VS VARIANCE TRADE OFF

Bias and Variance

The prediction errors in a model prediction includes

1. Bias
2. Variance

The understanding about these errors helps us to

❖Build an accurate model

❖And avoids overfitting and underfitting

Bias

Definition

Difference between predicted value and actual value

If the bias is too high , predicted value is faraway from the actual value

Variance

Definition

Variance is the change in prediction accuracy

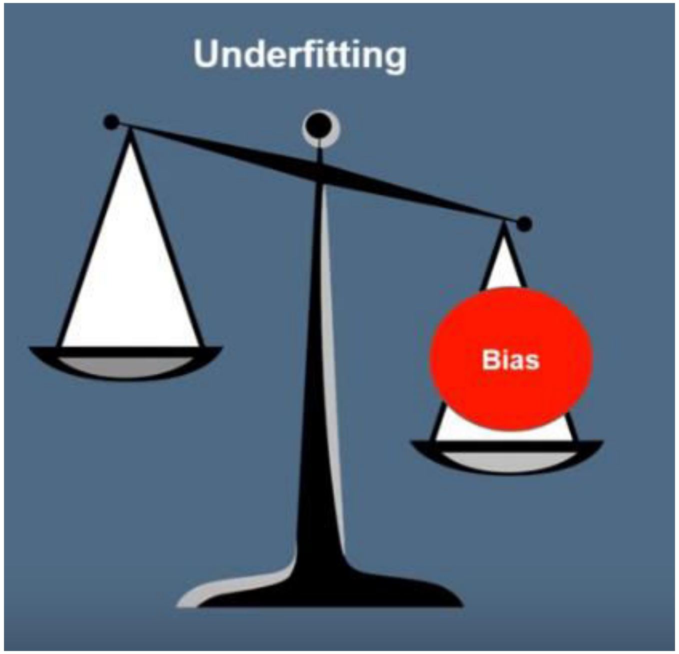
When the test data is not performed well as the train data, the model will vary

Bias Variance – Trade off

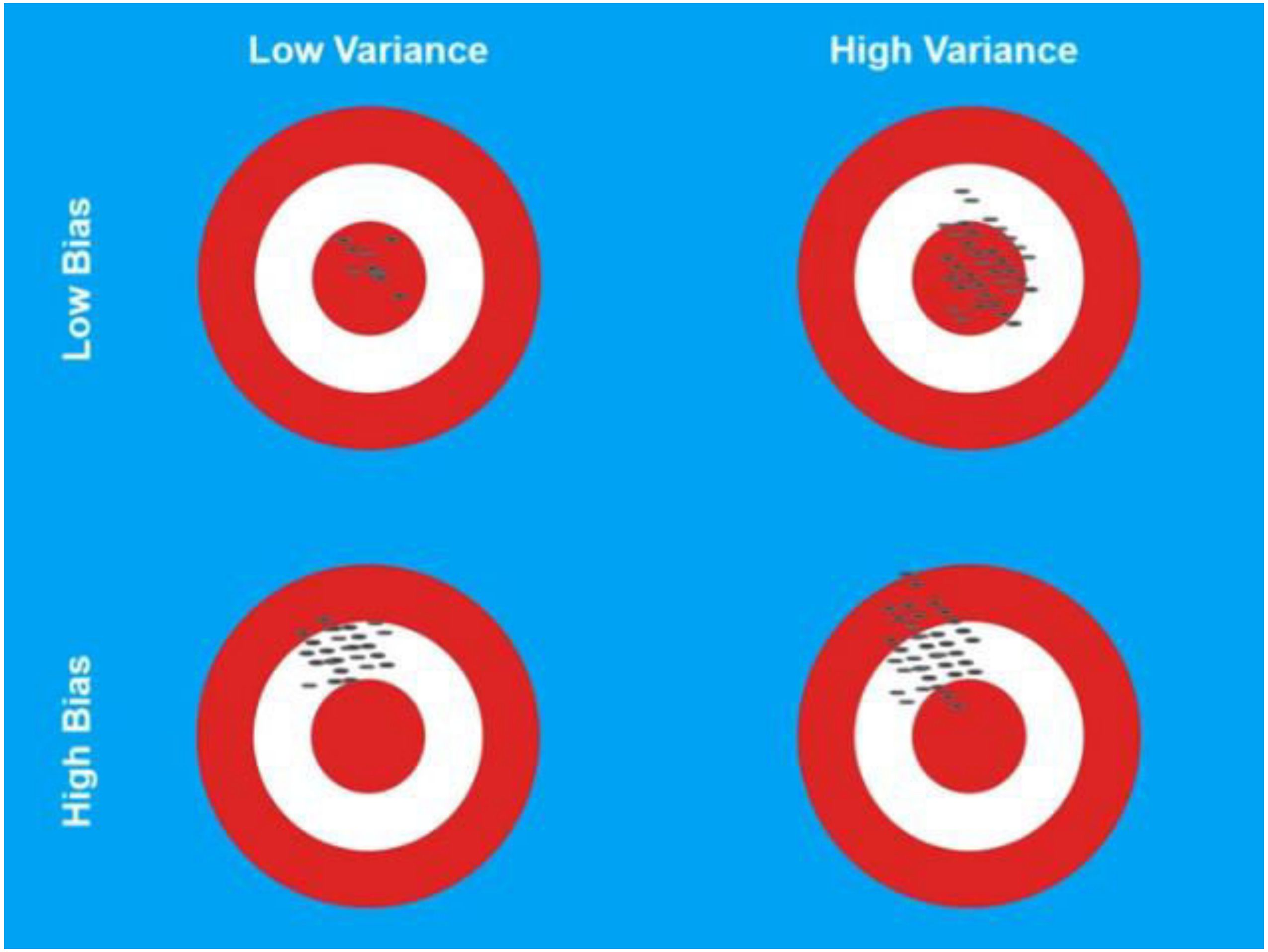
- If the model is too simple and has few parameters then it is having **High Bias and Low Variance**

- If the model is having large number of parameters then it is having **High Variance and Low Bias**

High Bias & High Variance



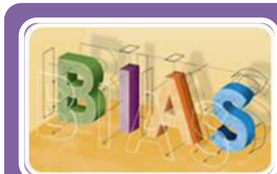
Scenarios



Summary

Difference between the predicted value and the actual value

Spread of data



High Bias - Occurs when a model not properly trained for training data which in turn has high error on both training and test data.

High Variance - Occurs when a model trains the training data much which in turn has high errors on test data.


Thank you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

CROSS VALIDATION

Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

Once we build a machine learning model, it has to be validated to ensure whether the model is working fine for the unseen data.

Helps us to predict the accuracy of our model.

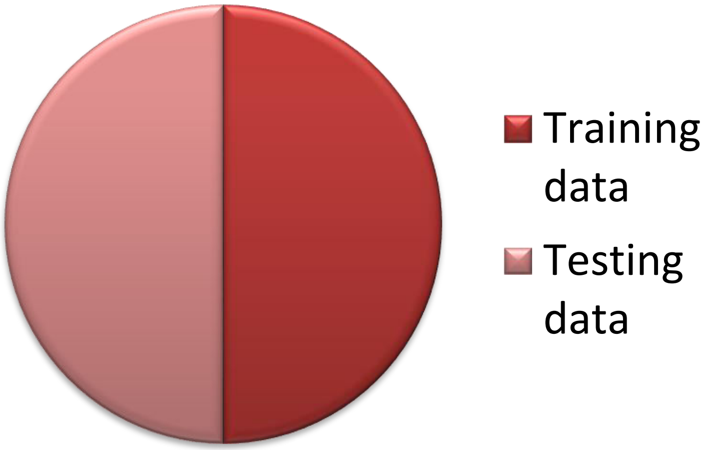
Steps Involved

1. Reserve some portion of sample data set.
2. Using the rest data set train the model.
3. Test the model using the reserve portion of the data set.

50% - 50%

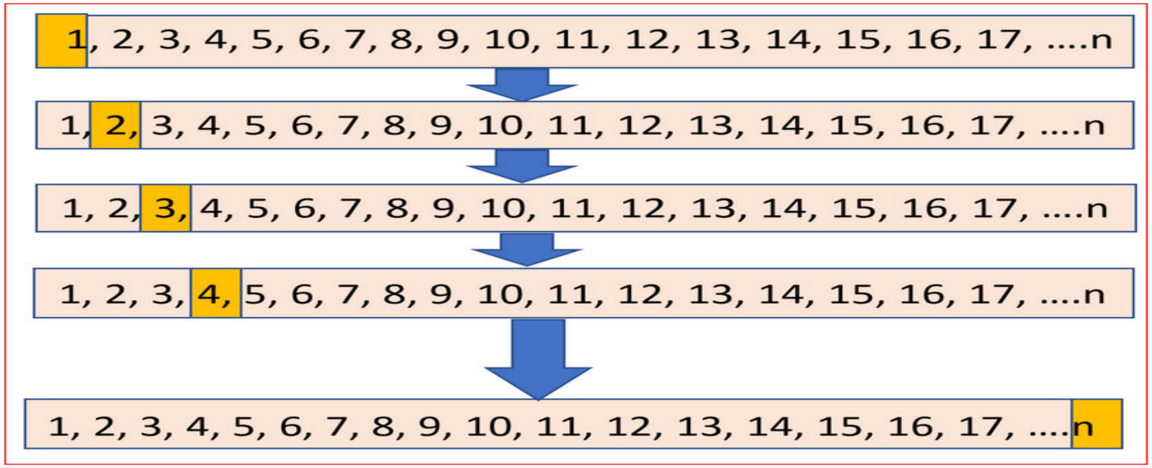
- In this method, we perform training on the 50% of the given data-set and rest 50% is used for the testing purpose.

- The major drawback of this method is that we perform training on the 50% of the dataset, it may possible that the remaining 50% of the data contains some important information which we are leaving while training our model i.e higher bias.



LOOCV (Leave One Out Cross Validation)

- Training on the whole data-set but leaves only one data-point of the available data-set and then iterates for each data-point.



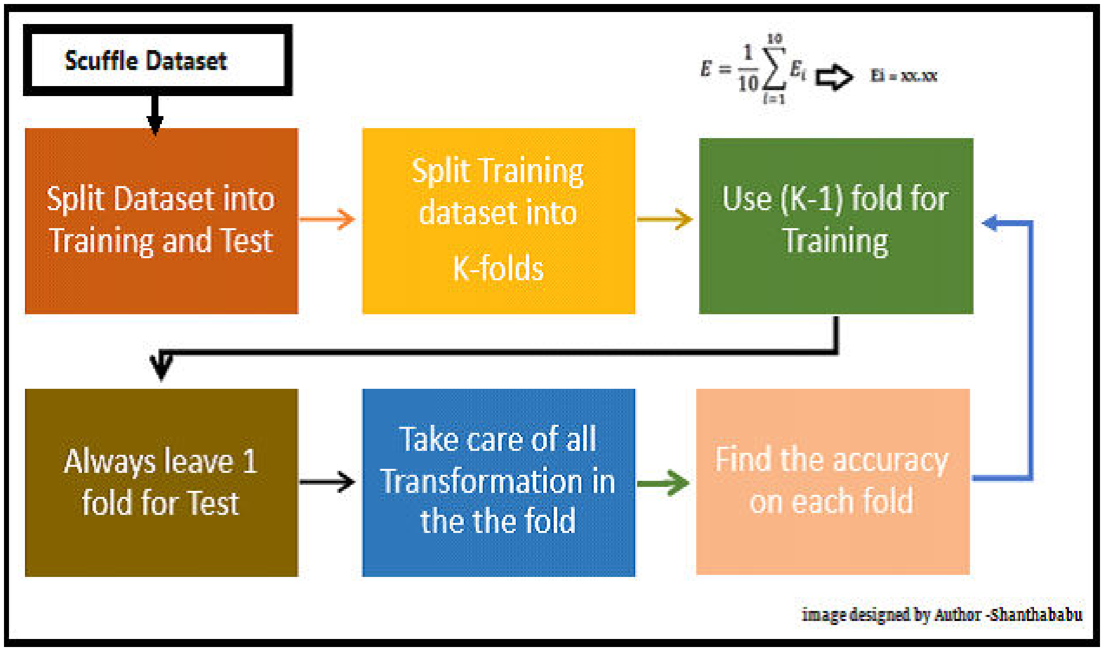
we make use of all data points and hence it is low bias.

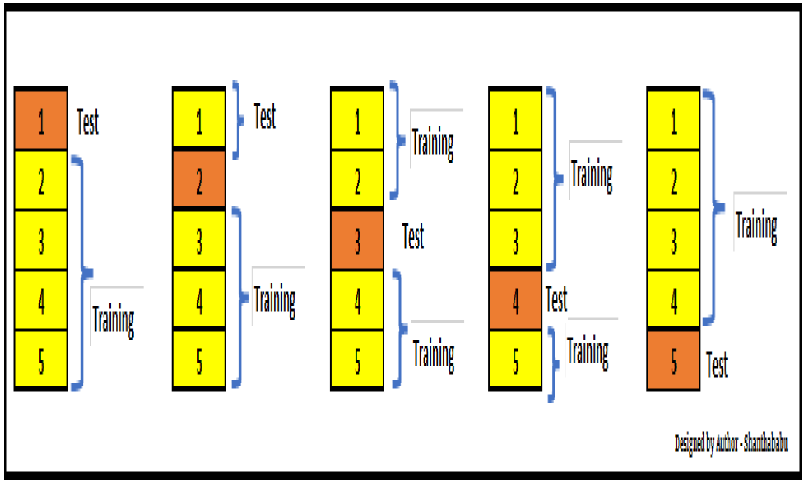
Higher variation in the testing model as we are testing against one data point

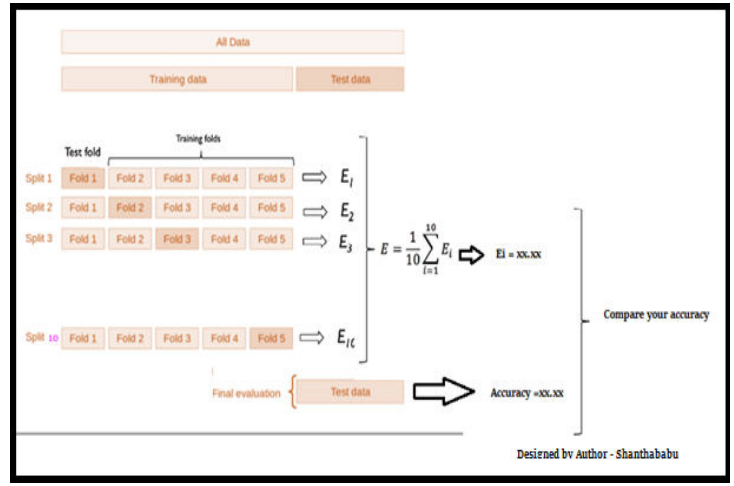
Lot of execution time

K-Fold Cross Validation

- ***Randomly dividing the dataset into k groups or folds*** of approximately equal size. The ***first fold is kept for testing*** and the **model is trained on k-1 folds**.
- Iterate k times with a different subset reserved for testing purpose each time.


Scuffle Dataset
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$
Ei = xx.xx
Split Dataset into Training and Test
Split Training dataset into K-folds
Use (K-1) fold for Training
Always leave 1 fold for Test
Take care of all Transformation in the the fold
Find the accuracy on each fold
image designed by Author -Shanthababu





K-Fold Cross Validation for the below purpose in the ML stream.
•Model selection
•Parameter tuning
•Feature selection

- Computation time is reduced as we repeated the process only 10 times when the value of k is 10.
- Reduced bias
- Every data points get to be tested exactly once and is used in training k-1 times
- The variance of the resulting estimate is reduced as k increases

The training algorithm is computationally intensive as the algorithm has to be rerun from scratch k times.


Thank you

18CSE392T – Machine Learning I

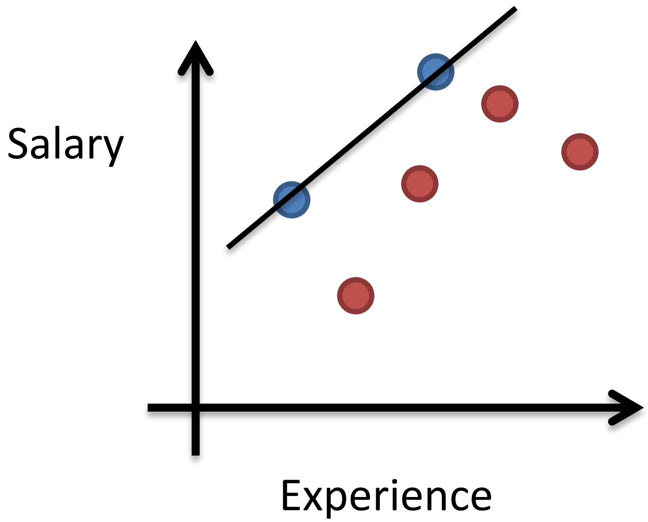
Department of Data Science and Business Systems

REGULARIZATION

Definition

Regularization is a technique used to reduce the error by fitting a model appropriately on a given training set and in turn helps to avoid over fitting

Linear Regression



Test data

Train data

y = mx + c

$$\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

- For training data - best fit - good results

(Low bias → Less error)

- For test data - Huge error

Overfitting

- w.r.t training data → Less error
- w.r.t test data → High error

Underfitting

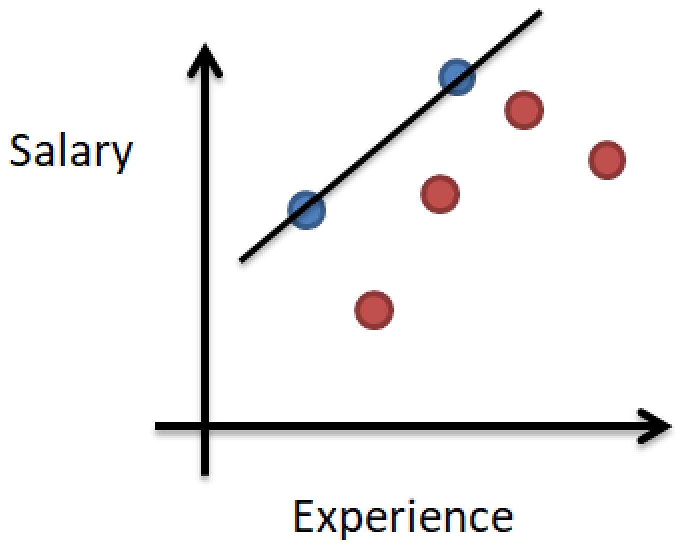
- w.r.t training data → High error
- w.r.t test data → High error

Types of Regularization

- L2 Ridge Regression
- L1 Lasso Regression

By using L1 and L2 regularization, we can convert High variance to Low variance

L2 - Ridge Regression



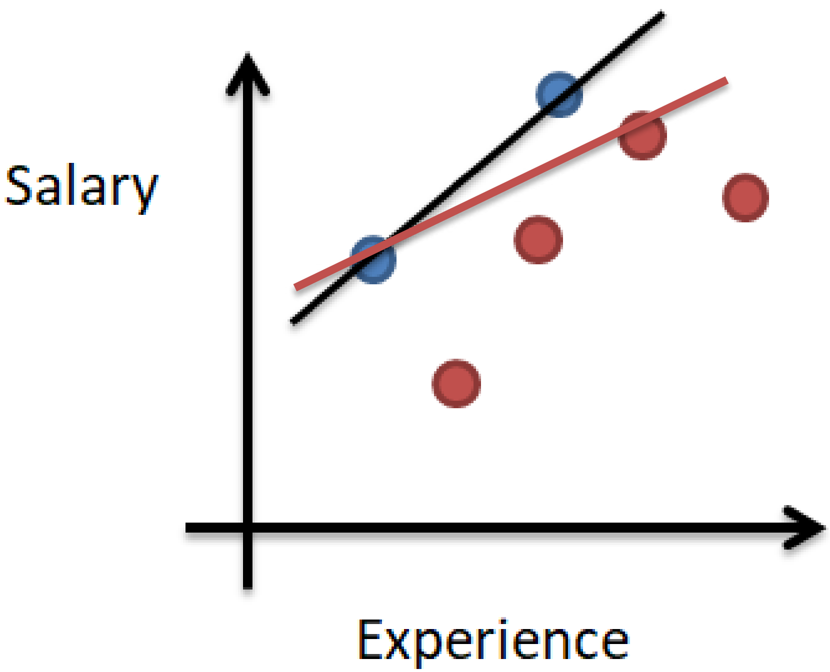
$$\text{Cost} = \underbrace{\sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2}_{\text{Loss function}}$$

L2 regularization on a dataset can be defined as follows:

$$\text{Cost} = \underbrace{\sum_{i=0}^{N}\left(y_i - \sum_{j=0}^{M} x_{ij} W_j\right)^2}_{\text{Loss function}} + \underbrace{\lambda \sum_{j=0}^{M} W_j^2}_{\text{Regularization Term}}$$

- W_j – slope

- Loss function is the traditional loss function where y is the dependent variable, x is the independent variables, and W is the kernel (weight matrices).
- The regularization term is added to the loss function. Note the regularization value is the sum of squared weight values across all the dimensions of a weight matrix.


Salary
Experience

L1 – Lasso Regression

- Not only for reducing the overfitting
- Also for feature selection (less slopes are removed – Extra features)

$$Y= m_1x_1 + m_2x_2+ m_3x_3 + m_4x_4 +...+m_nx_n +C$$

$$\alpha * | m_1 +m_2 +m_3 + ...+ m_n|$$

L1 – Lasso Regression

$$\widehat{y_i} = w_0 + \sum_{j=1}^{m} X_{ij} w_j$$

$$J(w) = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 + \alpha \sum_{j=1}^{m} |w_j|$$

Learning Curves

- A learning curve is a plot of model learning performance over experience or time.
- Learning curves of model performance on the train and validation datasets can be used to diagnose an underfit, overfit, or well-fit model

Cont..

- **Train Learning Curve**: Learning curve calculated from the training dataset that gives an idea of how well the model is learning.
- **Validation Learning Curve**: Learning curve calculated from a hold-out validation dataset that gives an idea of how well the model is generalizing.
- **Optimization Learning Curves**: Learning curves calculated on the metric by which the parameters of the model are being optimized, e.g. loss.
- **Performance Learning Curves**: Learning curves calculated on the metric by which the model will be evaluated and selected, e.g. accuracy.


Thank
you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

CLASSIFICATION

Definition

Classification is a task that makes use of machine learning algorithms that learns how to assign a class label to the data.

Types of Classification

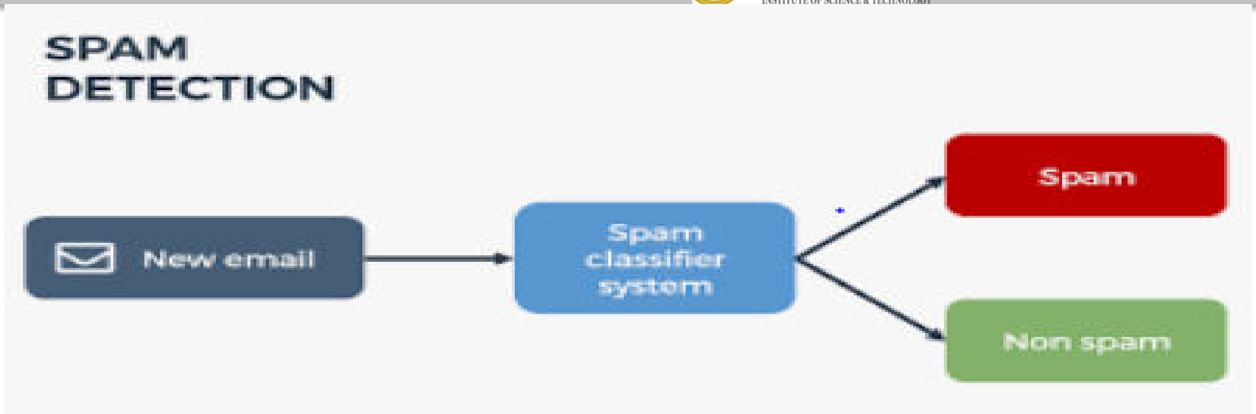
- Classification Predictive Modeling
- Binary Classification
- Multi-Class Classification
- Multi-Label Classification
- Imbalanced Classification

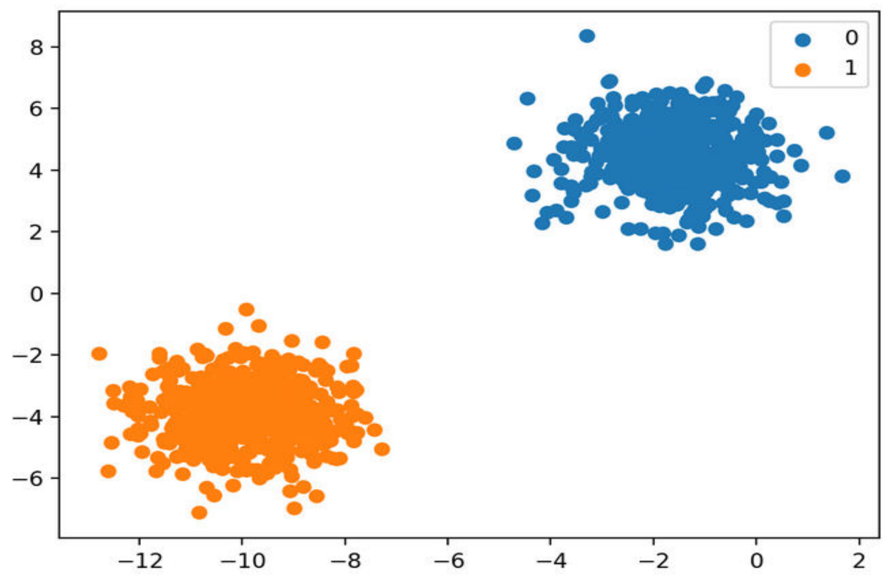
Classification Predictive Modeling

- Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data.

Binary Classification

- Binary classification refers to those classification tasks that have two class labels.
- Examples include:
  - Email spam detection (spam or not).
  - Churn prediction (churn or not).





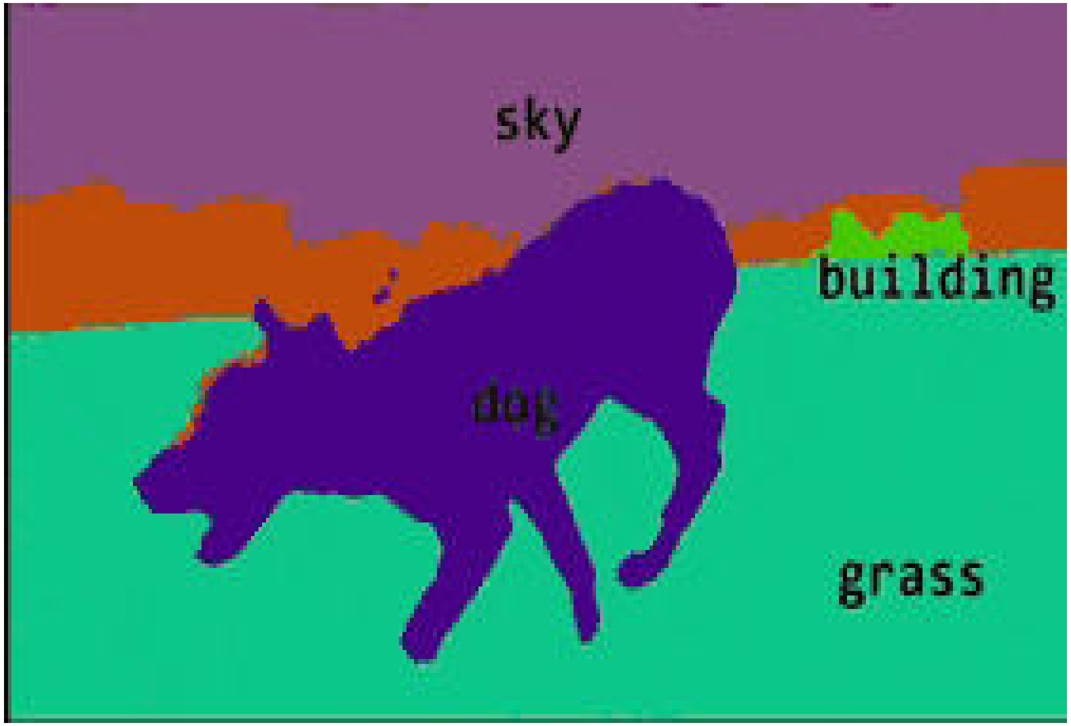
Scatter Plot of Binary Classification Dataset

Multi-class classification

- Multi-class classification refers to those classification tasks that have more than two class labels.
- Examples include:
  - Face classification.
  - Plant species classification.

Multi-label classification

- Multi-label classification refers to those classification tasks that have two or more class labels, where one or more class labels may be predicted for each example.
- Consider the example of photo classification, where a given photo may have multiple objects in the scene and a model may predict the presence of multiple known objects in the photo, such as "*bicycle*," "*apple*," "*person*," etc.
- This is unlike binary classification and multi-class classification, where a single class label is predicted for each example.


sky
building
dog
grass

Imbalanced Classification

- Imbalanced Classification refers to classification tasks where the number of examples in each class is unequally distributed.
- Mostly, imbalanced classification tasks are binary classification tasks where the majority of examples in the training dataset belong to the normal class and a minority of examples belong to the abnormal class.
  - Example - Fraud detection, Outlier detection.


Thank
you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

Agenda

- Error
- Noise
- Parametric Model
- Non-parametric model

ERROR

"how wrong is our estimation".

Error is a function which compares prediction of our model with real value

- An error measure is expressed as $\boldsymbol{E}(\boldsymbol{h}, \boldsymbol{f})$ (a hypothesis $\boldsymbol{h} \in \boldsymbol{H}$, and $\boldsymbol{f}$ is the target function).
- squared error: $\boldsymbol{e}(\boldsymbol{h}(x), \boldsymbol{f}(x)) = (\boldsymbol{h}(x) - \boldsymbol{f}(x))^2$
- binary error: $\boldsymbol{e}(\boldsymbol{h}(x), \boldsymbol{f}(x)) = [\![\boldsymbol{h}(x) \neq \boldsymbol{f}(x)]\!]$

NOISE

- Noise refers to irrelevant information in a dataset.
- Noise creates problem for machine learning algorithms because if it is not trained properly, the algorithms may think of noise as a pattern and can start generalizing from it, which is undesirable.
- We want the algorithm to make sense of the data and generalize the hidden properties of the data.

Collecting More data

The more data we collect, the better we will be able to identify the underlying phenomenon that is generating the data. This will ultimately helps in reducing the effect of noise.

Example – Survey in companies for mass scale

Principal Component Analysis(PCA)

PCA effectively reduces the dimension of the input data by projecting it along various axes.

For instance, consider projecting a point in a X-Y plane along X-axis.

This way, we are able to remove some noisy dimension along Y-Axis.

This process is referred as “dimensionality reduction”.

Regularization

- If the machine learning algorithm is flexible in learning more parameters which leads to overfits the noisy data.

- This can be avoided with the help of regularization.

Cross validation

- The hyper-parameters are tuned using the cross-validation data which is separate from the training data.
- This makes sure that the algorithm is able to avoid learning the noise present in the training data and rather generalize by a cross-validation procedure.

Parametric Model

- Parametric Algorithms has fixed number of parameters.
- It has faster computations and have strong assumptions about data.
- If the assumptions are right then algorithm works good whereas if the assumptions are wrong then algorithm works bad.
- Example: Linear Regression

Non – Parametric Model

- The non parametric algorithm uses flexible number of parameters.
- The number of parameters increases based on the learning from more data.
- This algorithm is slower in computation and makes few assumptions about data.
- Example: K-nearest neighbor


Thank
you

18CSE392T – Machine Learning I

Department of Data Science and Business Systems

LINEAR ALGEBRA

1. Notation

Knowing the notation will help you to read the algorithm descriptions in books, websites and papers to get an knowledge of the happenings.

2. **Operations**

The working in vectors and matrices can make things still clearer. This can be applied to descriptions and to code. We can apply simple operations like adding, multiplying, inverting, transposing, etc. on the matrices and vectors.

3. Matrix Factorization

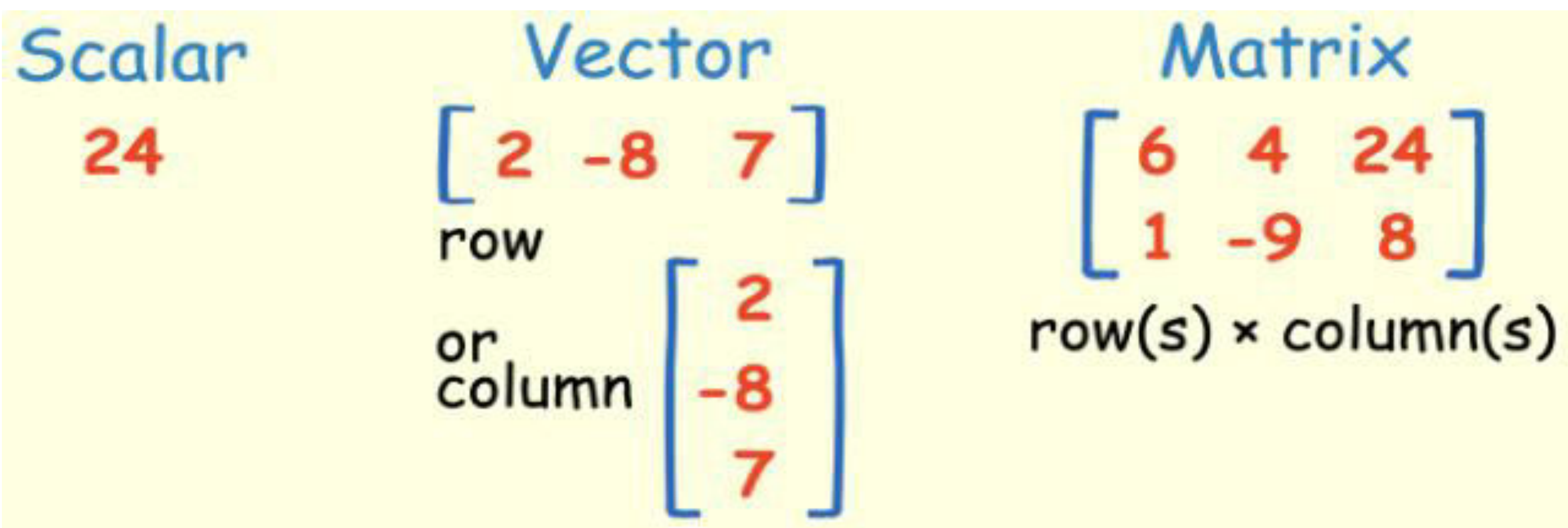
Matrix factorization algorithm works by decomposing the matrix into the product of two lower dimensionality rectangular matrices. The idea behind the matrix factorization is to represent users and items in lower dimensional latent space.

Notations

Symbol	Meaning
$\mathbf{A} \succ 0$	$\mathbf{A}$ is a positive definite matrix
$\mathrm{tr}(\mathbf{A})$	Trace of a matrix
$\det(\mathbf{A})$	Determinant of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	Determinant of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of a matrix
$\mathbf{A}^{\dagger}$	Pseudo-inverse of a matrix
$\mathbf{A}^{T}$	Transpose of a matrix
$\mathbf{a}^{T}$	Transpose of a vector
$\mathrm{diag}(\mathbf{a})$	Diagonal matrix made from vector $\mathbf{a}$
$\mathrm{diag}(\mathbf{A})$	Diagonal vector extracted from matrix $\mathbf{A}$
$\mathbf{I}$ or $\mathbf{I}_d$	Identity matrix of size $d \times d$ (ones on diagonal, zeros off)
$\mathbf{1}$ or $\mathbf{1}_d$	Vector of ones (of length d)
$\mathbf{0}$ or $\mathbf{0}_d$	Vector of zeros (of length d)
$\lvert\lvert\mathbf{x}\rvert\rvert = \lvert\lvert\mathbf{x}\rvert\rvert_2$	Euclidean or ℓ_2 norm $\sqrt{\sum_{j=1}^{d} x_j^2}$
$\lvert\lvert\mathbf{x}\rvert\rvert_1$	ℓ_1 norm $\sum_{j=1}^{d} \lvert x_j\rvert$
$\mathbf{A}_{:,j}$	j'th column of matrix
$\mathbf{A}_{i,:}$	transpose of i'th row of matrix (a column vector)
A_{ij}	Element (i,j) of matrix $\mathbf{A}$
$\mathbf{x} \otimes \mathbf{y}$	Tensor product of $\mathbf{x}$ and $\mathbf{y}$

Representation of Data

- In Linear Algebra, the data is represented by linear equations, which are presented in the form of matrices and vectors.



Scalar and Vector

- A scalar is simply a number.
- For example 45.
- A Vector is an ordered array of numbers.
- A vector can be in a row or a column.
- A Vector has a single index, which can point to specific value within the Vector.

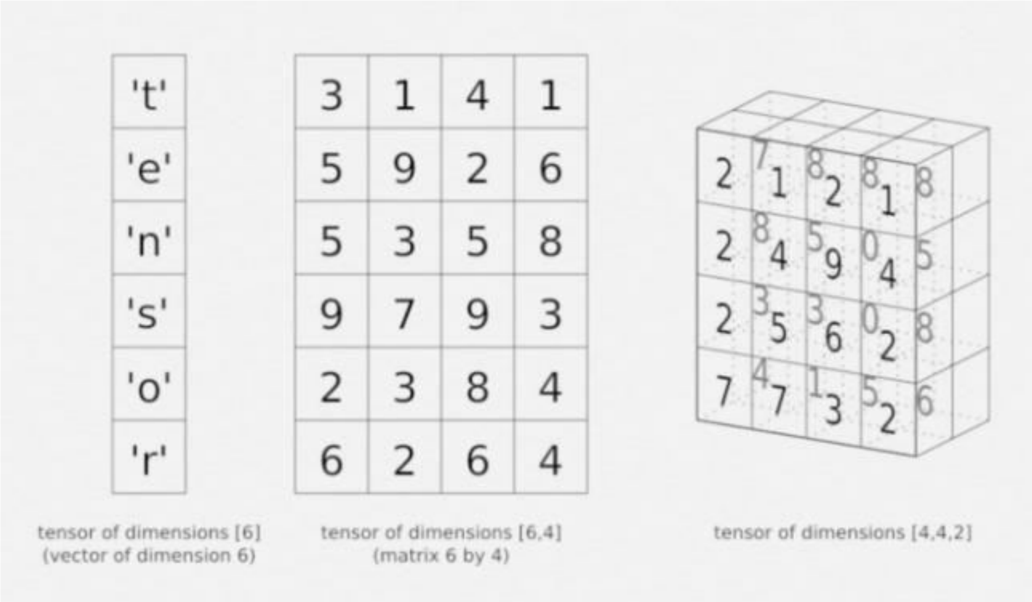
Matrix

- A Matrix is an ordered 2 dimensional array of numbers and it has two indices.
- The first one points to the row and the second one points to the column.
- Example: M21 refers to value in second row and first column.
- A matrix have multiple number of rows and columns.
- Vector is a matrix with one row or one column.

Tensor

- Tensor is an array of numbers arranged on a regular grid with a variable number of axes.
- A Tensor has three indices, the first one points to the row, the second to the column and the third to the axis.
- Example: T232 points to the second row, the third column, and the second axis.

Tensor



tensor of dimensions [6]
(vector of dimension 6)
tensor of dimensions [6,4]
(matrix 6 by 4)
tensor of dimensions [4,4,2]

Tensor

- Tensor is a multidimensional array.
- It can be a Vector and a Matrix, depending on the number of indices.
- A first-order Tensor would be a Vector (1 index).
- A second-order Tensor is a Matrix (2 indices).
- A Third-order Tensors (3 indices) and higher are called Higher-Order Tensors.

Operations

- **Matrix-Scalar Operations**
- **Matrix-Vector Multiplication**
- **Matrix-Matrix Addition and Subtraction**
- **Matrix-Matrix Multiplication**


Thank you