ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

UNIT III

Learning Objectives

Machine Learning:

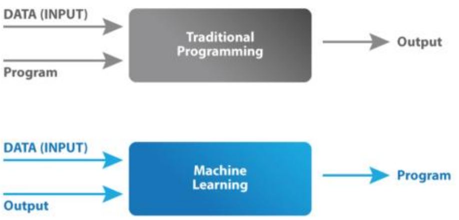
1. Introduction to Machine Learning
2. Why Machine Learning
3. Types of Machine Learning Problems
4. Applications of Machine Learning
5. Supervised Machine Learning: Regression and Classification
6. Regression: Simple, Multiple Regression, Least squares, Total Sum of Squares, Sum of Square of Residuals, Sum of Square of Regression, Odds, Odds ratio.
7. Classification: Logistic Regression
8. Accuracy Methods: Coefficient of determination, Correlation, confusion matrix, overfitting and underfitting, Bias and variance.

Machine Learning

- Machine learning is a field of **artificial intelligence** that allows systems to learn and improve from experience without being **explicitly programmed**.
- It is predicated on the notion that computers can learn from data, spot patterns, and **make judgments** with little assistance from humans.
- Good quality data is fed to the machines, and different algorithms are used to build ML models to train the machines on this data.
- The choice of algorithm depends on the **type of data** at hand and the type of activity that needs to be automated.

Why Machine Learning

- **How is it different from traditional programming?**



Types of Machine Learning Problems

1. Regression
2. Classification
3. Clustering
4. Time-Series Forecasting
5. Anomaly Detection
6. Ranking
7. Recommendation
8. Data Generation
9. Optimization

Regression

- When the need is to predict **numerical values**, such kinds of problems are called regression problems. For example, house price prediction, evaporation prediction, weather prediction etc.
- Algorithms utilized: Linear regression, K-NN, random forest, neural networks.

Classification

- When there is a need to **classify the data** in different classes, it is called a classification problem.
- If there are two classes, it is called a **binary classification** problem.
- When it is multiple classes, it is **multi-nomial classification.**
- Algorithms: Logistic regression, random forest, K-NN, gradient boosting classifier, neural networks

Clustering

- When there is a need to categorize the data points in **similar groupings** or clusters, this is called a clustering problem.
- Algorithms: Hierarchical clustering, Gaussian mixture models, K-Means.

Time Series Forecasting

- When there is a need to **predict a number based** on the time-series data, it is called a time-series forecasting problem.
- A time series is a sequence of numerical data points in successive order.
- Time series data means that data is in a series of particular time periods or intervals.
- Algorithms: ARIMA, SARIMA, LSTM, Exponential smoothing

Anomaly Detection

- When there is a need to find the outliers in the dataset, the problem is called an anomaly detection problem.
- If a given record can be classified as an outlier or unexpected event/item, this can be called an anomaly detection problem.
- Algorithms: Local outlier factor, One-class SVM

Ranking

- When there is a need to order the results of a request or a query based on some criteria, the problem is ranking problems.
- We rank the output of query execution based on scores we assign to each output based on some algorithms.
- Recommendation engines make use of the ranking algorithm to recommend the next items.
- Algorithms: Bipartite Rankboost, Bipartite RankSVM

Recommendation

- When there is a need to recommend such as "next item" to buy or "next video" to watch or "next song" to listen to, the problem is called a recommendation problem.
- The solutions to such problems are called recommender systems.
- Algorithms: Content-based and collaborative filtering machine learning methods.

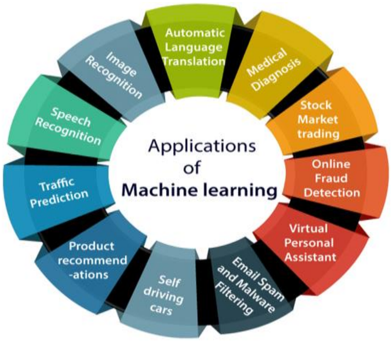
Data Generation

- When there is a need to generate data such as images, videos, articles, posts, etc., the problem is called a data generation problem.
- Generative adversarial network (GAN), Hidden Markov models.

Optimization

- When there is a need to generate a set of outputs that optimize outcomes related to some objective (objective function), the problem is called an objective function.
- Algorithms: Linear programming methods, genetic programming.

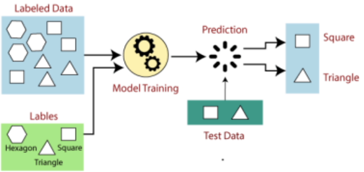
Applications of Machine Learning



Supervised Machine Learning

- Supervised learning is the types of machine learning in which machines are trained using well "**labelled**" training data, and on basis of that data, machines predict the output.
- Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.

Supervised Machine Learning (Contd.)



Supervised Machine Learning (Contd.)

- Steps involved in supervised learning algorithms:
  1. Collect the dataset
  2. Divide the dataset into training, testing and validation
  3. Determine input features
  4. Select suitable algorithms
  5. Execute the algorithm on the training dataset
  6. Evaluate the accuracy of the model by providing testing set.
  7. If model predicts correct output means, models is accurate.

Supervised Machine Learning (Contd.)

- There are broadly two types of supervised learning:
  - Regression
  - Classification
- Advantages of supervised learning:
  - Supervised techniques deal with labeled data where the output data patterns are known to the system.
  - Likely to produce more accurate outcomes.
- Disadvantages of supervised learning:
  - Training required lots of computation times.
  - Supervised learning cannot predict the correct output if the test data is different from the training dataset.

Regression

Age	Smoker	Incurred Cost
23	1	82
27	0	75
30	0	85
35	1	111
25	1	?

Regression (Contd.)

- Continuous variables
- What are regression problems?
- What are classification problems?
- Target Variables?
- Dependent Variables?
- Outcome Variables?
- Feature variables?
- Independent Variables?
- M(Example)?

Regression (Contd.)

- Regression Models:
  - Linear models
  - Non-linear models
  - Generalized additive models (GAMs)
  - Tree-based models
  - K- Nearest Neighbours
  - Decision Trees

Regression (Contd.)

- Linear Regression Models:
  - Model parameter
  - Bias parameter
  - Predicted Vs Actual

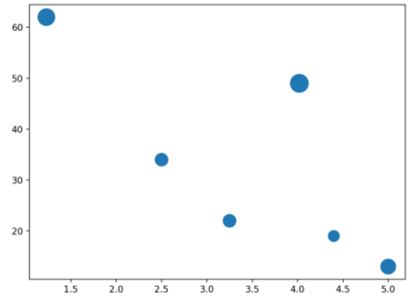
$$\underset{\text{Dependent Variable}}{Y_i} = \overset{\text{Constant/Intercept}}{\beta_0} + \underset{\text{Slope/Coefficient}}{\beta_1} \overset{\text{Independent Variable}}{X_i}$$

Regression (Contd.)

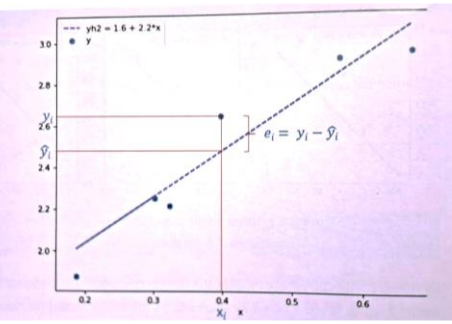
- Why these models are known as Regression Models?
- Why these models are known as Linear Regression Models?

Age	Smoker	Incurred Cost	Predicted Cost
23	1	82	
27	0	75	
30	0	85	
35	1	111	
25	1	?	

Regression (Contd.)



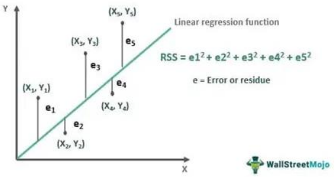
Regression- model error (Contd.)



Regression- model error (Contd.)

- Residual Sum of Squares (RSS)/Sum of Squared Errors (SSE)

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$



Determining Optimal Model Parameters Regression(Contd.)

- Regression Problem is to determine the **regression model parameters** that minimize the RSS metric over all the examples.
- In linear regression most frequently used technique is Stochastic Gradient Descent (**SGD**).
- Singular Value Decomposition (SVD).
- The process of finding these parameters are also known as training, building or fitting the model to the data.

Different Metrics Regression(Contd.)

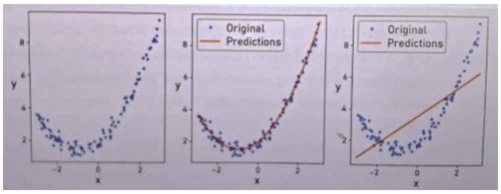
- RMSE
- TSS
- R-squared= 1-RSS/TSS

Non-Linear Models

- $S=1/2*a*t^2$
- Feature Engineering
- How do we know that we have to use a linear model or a non –linear model to model the relationship between features and target variables?
  - Plot the graphs
  - Use tree based techniques/ANN which automatically pick up the non-linearity if it exists.

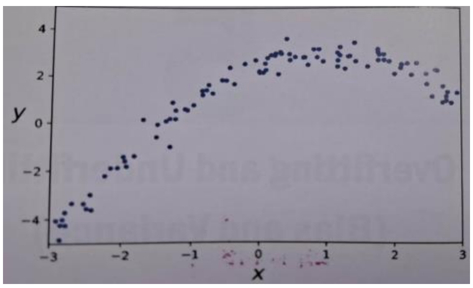
Sources of Model Error

- Measurement Errors
- Sub-optimal Model Parameters
- Sub-optimal Model
- Missing Features



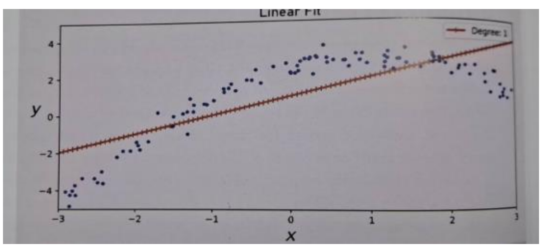
Overfitting and Underfitting

- $Y=2.5+X-0.5*X^2+noise$



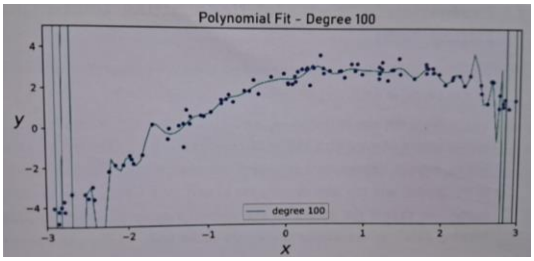
Overfitting and Underfitting

- $y = \beta_0 + \beta_1 * X$ (polynomial of degree 1)



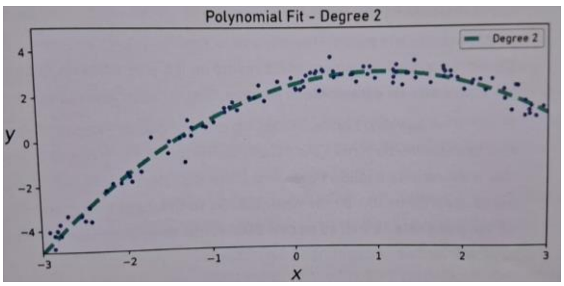
Overfitting and Underfitting

- $y = \beta_0 + \beta_1$*X+ β_2*X^2+….+ β_{100}*X^{100} (polynomial of degree 100)



Overfitting and Underfitting

- $y = \beta_0 + \beta_1$*X+ β_2*X^2 (polynomial of degree 2)



Model Validation (Train –Test Split)

- Training Set
- Testing Set

Classification

- In regression Target Variable (Dependent) takes continuous values, In classification the target variable takes **discrete values only**.
- This will requires us to use different algorithms and use different learning algorithms which works differently.
- A variable is called **categorical**, if it can take any value from a fixed set of values.
- The values of a categorical variable are said to be discrete.

Classification (Contd.)

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	diagnosis
16.13	20.68	108.10	798.8	0.11700	M
14.92	14.93	96.45	686.9	0.08098	B
13.75	23.77	88.54	590.0	0.08043	B
13.40	20.52	88.64	556.7	0.11060	M
20.29	14.34	135.10	1297.0	0.10030	M

Classification (Contd.)

- ML algorithms typically works with **numerical data**, Hence the categorical variables are converted into numbers before they are passed on to any ML algorithm.
- Here in the previous slides the case of M and B category could be represented as B→ 0, M→1.
- These are mapped values only, refrain from thinking of ordering when dealing with these mapped values.

Classification (Contd.)

- This is to be noted that only **target variable** in classification problems are categorical or discrete.
- The **independent variables** or features **may be continuous** and need not be discrete.
- Thus, Classification problems needs to be addressed with different ML algorithms.

Approaches followed by Classification

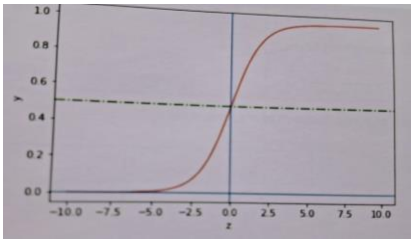
- Classification usually work by predicting the probability of the example belonging to one of the possible classes that the dataset has.
- Class probability are subsequently converted into class labels.
- As an example the target variables has values ['pink', 'black','white'] the class prob. For an example can be [0.3,0.25,0.45]

Approaches followed by Classification

- In case of two-class classification problems, note that the algorithm is expected to generate two prob. P_0, $P_{1.}$
- Since P_0+ $P_{1=}$ 1
- $P_0 = 1 - P_1$

Logistic Regression

- Model used for classification problems.
- Sigmoid function is used in this type of problems.



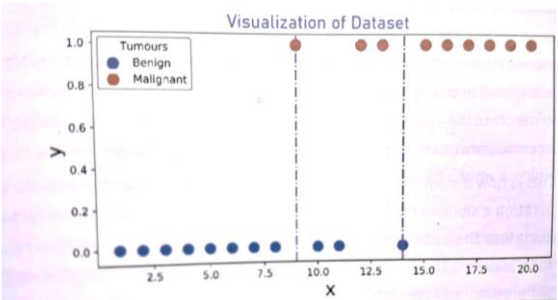
Logistic Regression

- For a two class classification problem, the classifier needs to generate a probability p1 only for the class 1, the probability for class 0 is simply p0=1-p1.
- The probability lies between 0 to 1.
- The sigmoid function values lies between 0 to 1.
- Hence, can be used a prob.

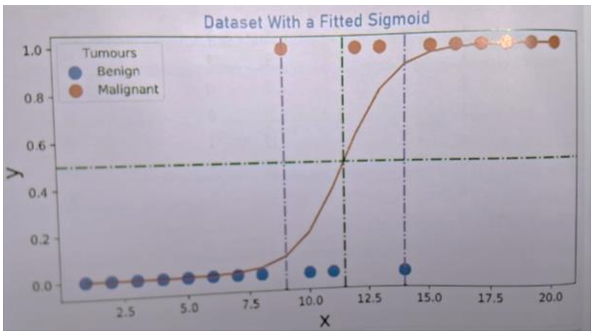
Logistic Regression

x	y	x	y	x	y
1	0	9	1	15	1
2	0	10	0	16	1
3	0	11	0	17	1
4	0	12	1	18	1
5	0	13	1	19	1
6	0	14	0	20	1
7	0				
8	0				

Logistic Regression



Logistic Regression



Logistic Regression

X	y	p_0	p_1	p_2	p_3	$\hat{y}$
$features^0$	$y^0 = 2$	$p^0_0 = 0.2$	$p^0_1 = 0.1$	$p^0_2 = 0.25$	$p^0_3 = 0.45$	$\hat{y}^0 = 3$
$features^1$	$y^1 = 1$	$p^1_0 = 0.35$	$p^1_1 = 0.4$	$p^1_2 = 0.15$	$p^1_3 = 0.1$	$\hat{y}^1 = 1$
$features^2$	$y^2 = 0$	$p^2_0 = 0.26$	$p^2_1 = 0.11$	$p^2_2 = 0.36$	$p^2_3 = 0.27$	$\hat{y}^2 = 2$
$features^3$	$y^3 = 3$	$p^3_0 = 0.15$	$p^3_1 = 0.18$	$p^3_2 = 0.28$	$p^3_3 = 0.39$	$\hat{y}^3 = 3$

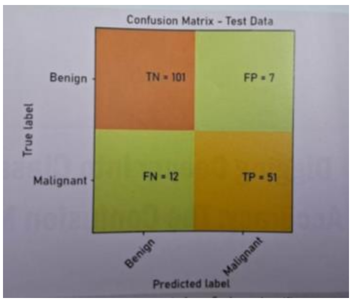
predicted probabilities (columns p_0–p_3) derived label (column $\hat{y}$)

Logistic Regression

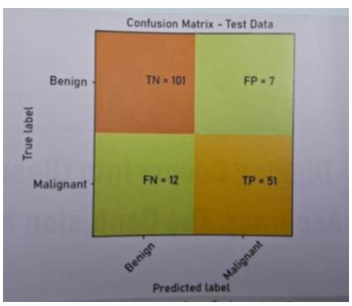
- CROSS entropy matric is used instead of RSS.
- Loss=y_{act} $\ln(y_{pred})$-(1- y_{act})ln(1- y_{pred})

Accuracy Methods

- Accuracy=n/m



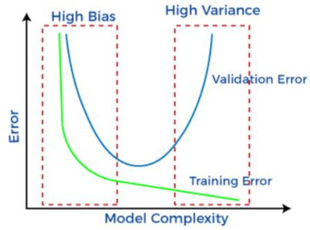
Confusion Matrix



- Accuracy=TP+TN/TP +TN+FP+FN
- Accuracy (negative class) also know as sensitivity=TN/TN+FP
- Accuracy(positive class)or recall or specificity= TP/TP+FN
- Precision=TP/TP+FP

Bias

- Bias and Variance help us in parameter tuning and deciding better-fitted models among several built.



Bias (Contd.)

- While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.
- **Low Bias:** A low bias model will make fewer assumptions about the form of the target function.
- **High Bias:** A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. **A high bias model also cannot perform well on new data.**

Bias (Contd.)

- Generally, a linear algorithm has a high bias, as it makes them learn fast. The simpler the algorithm, the higher the bias it has likely to be introduced. Whereas a nonlinear algorithm often has low bias.
- Some examples of machine learning algorithms with low bias **are Decision Trees, k-Nearest Neighbours and Support Vector Machines**. At the same time, an algorithm with high bias is **Linear Regression, Linear Discriminant Analysis and Logistic Regression.**

Bias (Contd.)

Ways to reduce High Bias:

- High bias mainly occurs due to a much simple model. Below are some ways to reduce the high bias:
- Increase the input features as the model is underfitted.
- Decrease the regularization term.
- Use more complex models, such as including some polynomial features.

Variance

Variance tells that how much a random variable is different from its expected value.

Ideally, a model should not vary too much from one training dataset to another, which means the algorithm should be good in understanding the hidden mapping between inputs and output variables.

Variance errors are either of low variance or high variance.

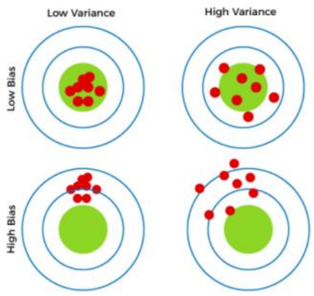
Some examples of machine learning algorithms with low variance are, **Linear Regression, Logistic Regression, and Linear discriminant analysis**. At the same time, algorithms with high variance are **decision tree, Support Vector Machine, and K-nearest neighbours.**

Variance (Contd.)

Ways to Reduce High Variance:

- Reduce the input features or number of parameters as a model is overfitted.
- Do not use a much complex model.
- Increase the training data.
- Increase the Regularization term.

Variance (Contd.)



Variance (Contd.)

Low-Bias, Low-Variance:

- The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.

Low-Bias, High-Variance:

- With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an overfitting

Variance (Contd.)

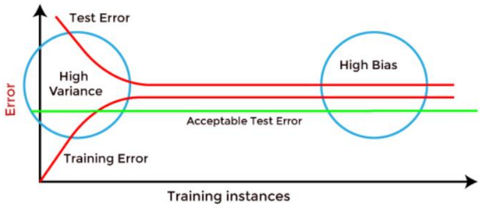
High-Bias, Low-Variance:

- With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to underfitting problems in the model.

High-Bias, High-Variance:

- With high bias and high variance, predictions are inconsistent and also inaccurate on average.

Variance (Contd.)



Variance (Contd.)