Day-50

Date 24/5/2022

Machine Learning

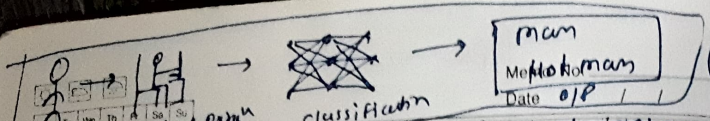
1. Introduction

Data science :

- First of all Data science is all about data only.
- It is a process of extracting knowledge and insights from data by using few methods or scientific methods.
- Data science is domain of study that concerns daily with vast no of data only using some modern tools & technique to find some unseen patterns, to derive some meaningful information to make business decisions. To make such decisions on the base of data it uses complex machine learning algorithms to build the predictive model.

e.g Fruits :

orange >> orange, circular, sweet
chikoo >> brown, oval, sweet
orangl >> orange, circular, bi sour
chikoo >> brown, oval, bi sour



1. Machine Learning : ML is a technique to implement AI that can learn from the data by themselves without being explicitly programmed.

- It is a branch of AI and computer science which focuses on the use of data & algorithms to imitate the way that human learn, gradually improving it's accuracy.

- It is a study of computer algorithms that improve automatically through experience and use of data.

2. Deep learning DL is a subfield of ML that uses Artificial Neural Networks to learn from the data

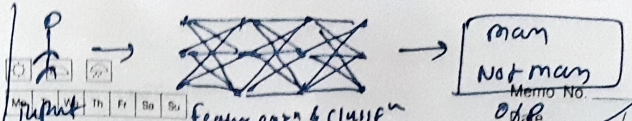
- Deep learning is a subset of a machine learning, which is essentially a neural network with three or more layers.

- It uses for many AI applications and service that improve automation performing some analytical and physical tasks without human intervention.

- Deep learning technology lies behind so many everyday products and services such as, digital assistants, voice enabled TV remotes.

3. Natural Language processing:

- It refers to the branch of AI again, which



concerned with giving computers the ability to understand text or spoken words in such a way human beings can.

4. Statistics

- This branch of applied mathmatics that involves the collections, descriptions, analysis and inference of conclusion from the data.

6. Data Visualization

- Data visualization is a graphical representation of information and data. By using elements like charts, pie charts, graphs, maps, data visualization tools provide an accessible way to see and understand trends, outliers, patterns in data.

Data:

- ~~Structured~~ : CSV, EXCEL
- ~~Semistructure~~ : Json, Html
- ~~Unstructured~~ : text, images, audios, videos

#1. Machine Learning

Data ~~sets~~.

↳ (1) Training dataset (> 70%) 70% data for training
(2) Testing dataset (30% <) unseen data.

Memo No
Date / /

e.g Fruit understanding machine.

Fruit >>	colour	shape	taste
orange >>	orange,	circular,	sweet
chikoo >>	brown,	oval,	sweet
chikoo >>	brown,	oval,	lil sour
orange >>	orange,	circular,	sweet
orange >>	orange,	circular,	lil sour
chikoo >>	brown,	oval,	lil sour
chikoo >>	brown,	oval,	sweet
orange >>	orange,	circular,	sweet

Application of ML

1. Chatbot (voice based, text based)
2. Recommendation system (flipkart, amazon, netflix)
3. Medical field (cancer, covid)
4. Finance domain (bank)
5. Insurance Domain
6. Stock market
7. Self driving car
8. House price prediction
9. Object detection
10. Face recognization
11. Weather forecasting
12. Speech recognization

Memo No.
Date / /

13. Cyber security.
14. Google maps.
15. human activity recognization.

Types of Machine Learning

① Supervised ML — The ML learning algorithm from labelled data

② Unsupervised ML — unlabelled data

③ Reinforcement ML — It is an area of ML concerned with how intelligent agents take actions in an environment to maximize its rewards.

(x) Independent Variable >> which are given as input for prediction, predictors.

(y) Dependent Variable >> Target variable, Outcome Variable.

training data >> Independent + Dependent

Testing data >> Independent

e.g. unsupervised ML BAA ABA BBB A →  → AAAAA BBBBB cluster

Two types of unsupervised ML

1. Clustering — (grouping the similar data points).

- It finds the similarity in between the data points and according it make clusters.

2. Association (find important relationship betn data points)

- It finds the relationship between data points and then make the groups out of it.

1. K means clustering >> distance base
2. Hirarchical clustering
3. PCA >> principle component analysis
4. ~~KNN (K-Nearest neighbors)~~

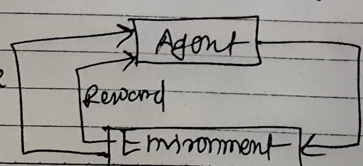
Unsupervise ML Algorithm

3. Reinforcement Machine Learning

Rewards based learning / Feedback
Appriciate / punish

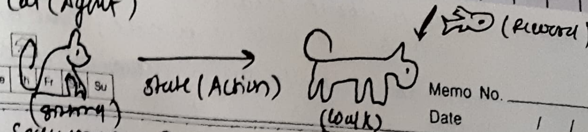
- Reinforcement ML is a type of machine learning in which agent explores an environment to learn how to perform desired tasks by taking actions with good outcomes and avoiding actions with bad outcomes.

e.g ①



① PK movie (Planet Person) → Gandhi image on Note.

③ Cat (Agent)



- A sequence of successful outcome will be reinforced to develop the best recommendation or policy for a given problem.

Steps of Data Science project

1. Problem statement (what does he required) client / depends on client requirement.
2. Data Gathering >> Database, CSV, Excel, Text, images, videos, audios.
3. Exploratory Data Analysis: (File Analysing as per categorical / numerical, words) pandas, matplotlib, seaborn.
   - 3.1 Univariate analysis >> Analysing a single feature at a time.
   - 3.2 Bivariate analysis >> We analyse two different features to check their relationship.
4. Data preprocessing / Feature Engineering: (Null values, Unwanted things to be cleaned)
   - Handling outliers.
   - Normalization
   - Standardization >>.
     20-70 >> 40000-100000 40x
     one hot encoding, label encoding, get dummies, ~~feature selection~~
5. Feature selection (model building करने से पहले)
   - 5.1 Fisher's score
   - 5.2 Forward feature selection
   - 5.3 Backward feature selection
   - 5.4 Correlation
   - 5.5 Random Forest feature selection.

Classification :- It is about predicting a class or discrete values e.g male or Female ; True or False.

Regression :- It is about predicting a quantity or continuous values e.g salary, age, price.

Date / /

6. Model Training : (ML algorithm)
   - 6.1 Linear regression
   - 6.2 Logistic regression
   - 6.3 Knn
   - 6.4 DT

7 - Model Evaluation : (Algorithm on test data)

(Tay [illegible]) Classification > confusion matrix, classification report, accuracy score.

(Value [illegible]) Regression > MSE, R2 Score

8. Deployment : (AWS, GCP, Azure)

Things to focus in each & every algorithms

1. Concept.
2. Assumption
3. Effect of missing values & and outliers.
4. Advantages
5. Disadvantages
6. Applications.

Day-52

Always target variable is only one

Memo No.

Date 26 / 5 / 2022

VIMP 1. Linear Regression Algorithm

- It is type of supervised ML algorithm.
- This is a very basic algorithm.
- It is the algorithm which highly used initially and even highly studied as well.

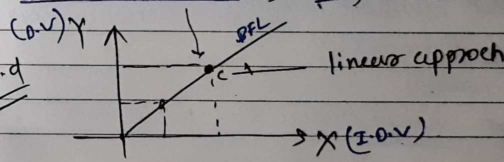
Why name is Linear

- Linear regression is a linear approach to modelling the relationship between dependent and independent variable.

Screen sharing => if not able to explain orally.

- Main target of linear regression is to find out the best fit line. & look for linear regression with D.V & I.D.V

e.g



Day-52

Types of Linear Regression Algorithm

1.1 Simple linear regression. { one D.V & one I.D.V
   - 1 Dependent variable
   - 1 independent variable

(SRL) Target variable cont.

1.2. multiple linear regression >> (1 से ज्यादा variable I.D.V)

one dependent variable, more than one independent variable.

1 → D.V

>1 → IDV

- Target variable should be continuous in nature.

(Y) Outcome / Dependent Variable — Target variable / outcome variable

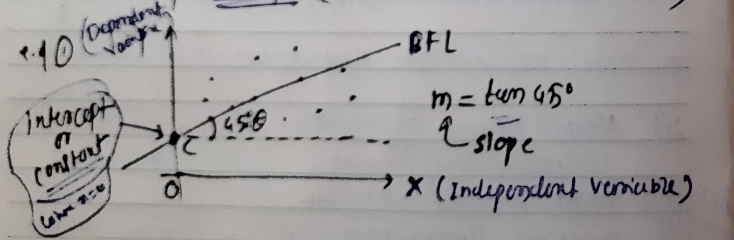
Memo No.
Date / /

- Linear regression is an algorithm where we basically try to understand the linear relationship b/w (dependent variable and one or more independent variable.

- It is based on line
So when it comes to linear regression we basically have a straight line. and straight line is nothing but linear equation. that is why this algorithm is known as Linear regression.

- equation of line => $y = ax + b$ or $y = mx + c$ ← simple linear regression eqn.

y = dependent variable / outcome variable
x = independent variable
m = slope / position
c = intercept (where line of y when x is zero)



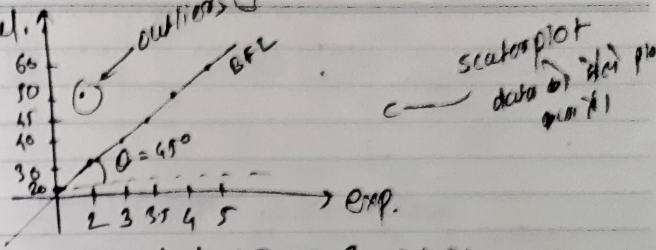
e.g. Experience Salary

2	20K
3	40K
3.5	45K
4	50K
2	50K >> outliers
5	60K

exp — independent variable.
Sal — dependent variable

(intercept => where line of y when x is zero)

Memo No.
Date / /



scatterplot → data ko plot karna

If we want to find out 4.5 => ? salary.

Simple Linear Regression → $y = mx + c$

$mx + c = 1 \times 4.5 + 20K$ ∵ tan 45 = 1

$= 20004.5$

4.5 exp >> 20004.5 (50K से ज्यादा सैलरी मिलेगी)

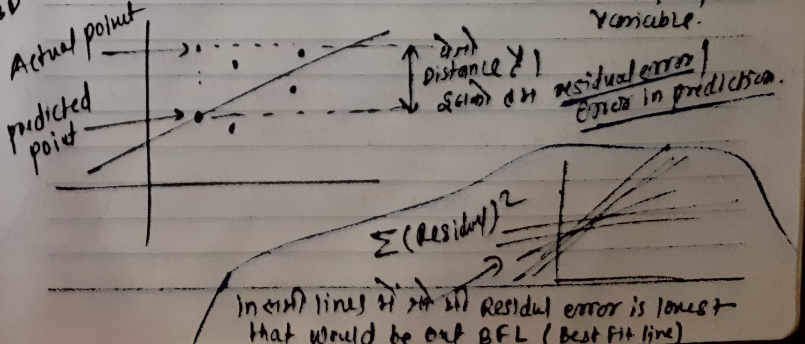
multiple linear regression

Suppose we have two independent variable.

2D multiple Linear → $y = m_1x_1 + m_2x_2 + c$

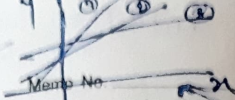
multiple Linear 3D → $y = m_1x_1 + m_2x_2 + m_3x_3 \ldots\ldots + m_nx_n + c$

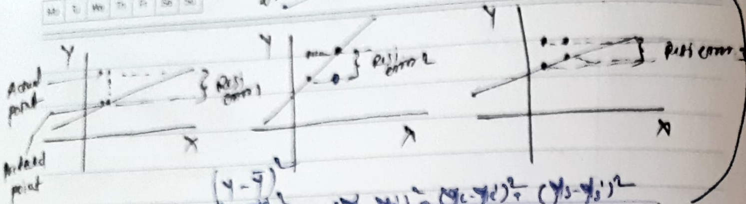
n = no. of independent variable.



Residual error or error in prediction

Memo No.
Date / /





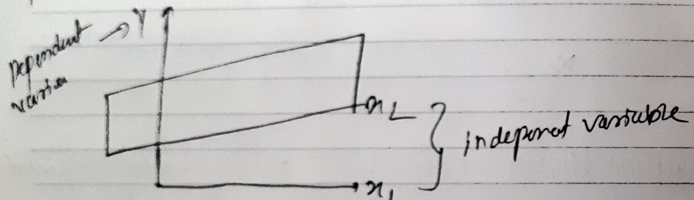
$(y-\bar{y})^2$

$(y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + (y_3-\hat{y}_3)^2$

$\Sigma(\text{Resi})^2 = (\text{Resi } 1)^2 + (\text{Resi } 2)^2 + (\text{Resi } 3)^2$

- इसमें से जो value कम है वो होगी है हमारी BFL.

eq ① ⇒ 3D गावा से ज्यादा 2D to 3D imagine कर सकते



e.g for 1 से ज्यादा example.

- The moto of linear regression is we need Linear data.

Imp Assumption

Independent variable supposed to be independent with each other. It simply means they should not have any enfluence upon each other.

imp assumption for L.R.

eg. x_1, x_2 >> There should not be any relation betn both of them.

Eg. Car. >> Weight, HP, Mile/gallon etc.

Memo No.
Date / /

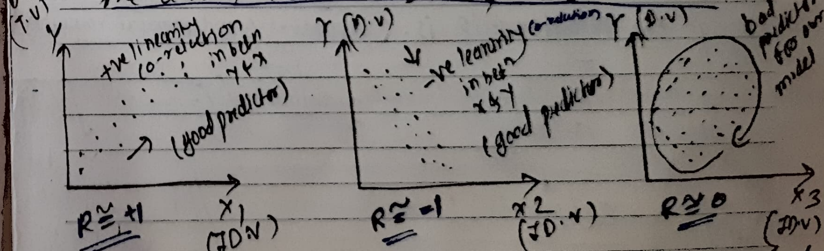
in real world more the wt, more the HP but our model always consider that there should be no relation in between this two variable. This problem is known as Co-linearity.

$y = m_1x_1 + m_2x_2 + c$

$m_1 \Rightarrow m_1$ is how much y changes whenever x_1 changes by one unit keeping x_2 constant.

$m_2 \Rightarrow m_2$ is how much y changes whenever x_2 changes by one unit keeping x_1 constant.

x_1 and x_2 (independent variable) >> if this two variables does have some relation or they are not independent on each other, then it becomes difficult understand the actual relation bet dependent variable.



- There should be high co-relation in betn dependent variable & independent variable.

- There should not be co-relation in betn independent variable

co-efficient of correlation in betn D.V & ID.V = R (-1 to +1)

Memo No. ______ Date / /

$y = m_1x_1 + m_2x_2 + m_3x_3 + c$

x_1 >> positively linear/upward >> good predictor

x_2 >> Negatively linear/downward >> good predictor

x_3 >> non-linear/not having any co-relation with dependent variable >> bad predictor.

Q* How can we measure the strength of the relationship betn independent variable and target/dependent variable?

↳ For this we do have one metrix which is called as Coefficient of co-relation. which is denoted by R. Range of R is always in between -1 to +1. It can not be beyond this range.

[0.3 to 0.7 (worst)]

R value is an indicator for understand that how strong the relationship is. (dependent & independent variable)

$$R = \Sigma(x_i - \bar{x}\text{ bar/mean})(y_i - y\text{ mean/bar}) \Big/ \sqrt{\Sigma(x_i-\bar{x})^2}\sqrt{\Sigma(y_i-\bar{y})^2}$$

coefficient of i.e co-relation (ID.V, D.V)

$$R = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2}\sqrt{\Sigma(y_i-\bar{y})^2}}$$

This is main formula

$$R = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2}}$$

R = coefficient of co-relation

x_i = values of x variable in sample

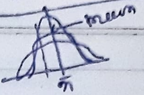
$\bar{x}$ = mean of values of x-variable

y_i = values of y variable in sample

$\bar{y}$ = mean of values of y variable

Cov ⇒ Co-variance

Variance



Variance refers to the spread of a data set around it's mean value i.e.($\bar{x}$)

- Variance is measurement of the spread between nos. in data set (or) we can also say variance measures variability.

$\bar{x}$ → mean value

$$\sigma^2\ Var = \sum_{i=1}^{i=n}\frac{(x_i-\bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

Here we have checked co-relation of R

Covariance is directional relationship betn two random variables.

$$Cov = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}$$

cov(x,y)

n-1 population covariance

n - sample covariance

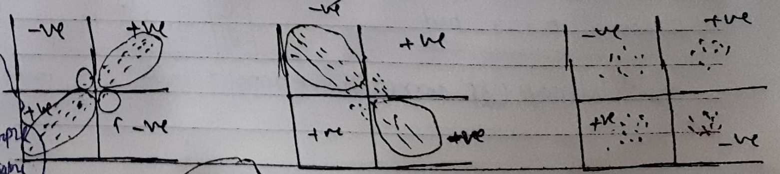
$$R = r_{xy} = \frac{Cov}{\sigma_x\ \sigma_y}$$

$\left(\sqrt{\Sigma(x_i-\bar{x})^2}\right)$ $\sqrt{\Sigma(y_i-\bar{y})^2}$

-1 to +1

we call R r value as a co-efficient as it is simply a ratio.

- There should be high co-variance between dependent and independent variable & zero co-variance betn two independent variables. (having zero co-variance is an ideal situation)



- more data point in +ve quad. $R \simeq +1$

more data point in -ve $R \simeq -1$

+ve = -ve data point. $R \simeq 0$

Kle are check correlation of iris dataset

Day-53

Memo No.

Date 27/05/2022

e.g. In jupyter notebook — file co-relation.

We are cheking a iris data basic idea

```
import numpy as np
import pandas as pd

df = pd.read_csv('Iris.csv')
df.head()   # [illegible] colowm drop [illegible]
df.drop(['ID', 'species'], axis=1, inplace=True)
df.head()
```

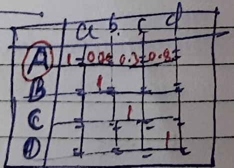
axis = 0 rows
axis = 1 columns

(coefficient of corelation)

Now we are cheking co-reation [illegible]

```
import seaborn as sns
import matplotlib.pyplot as plt
df.head()   # required colowm [illegible]
sns.pairplot(df)
df.corr()   # coefficient of co-relation.
```

- R approx -1 >> good
- R approx +1 >> good
- R approx 0 >> bad

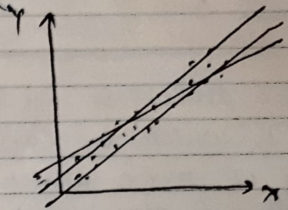


```
sns.heatmap(df.corr(), annot=True)   # graph visualize
```

Memo No.

Date / /

Good predector (Gradient Decent Algorithm)

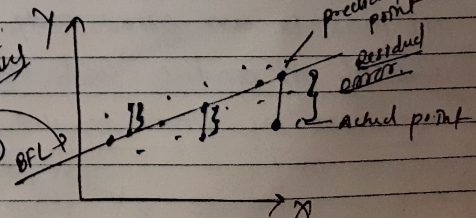


Data point the linear w.r.t Y

$y = mx + c$ (internally algorithm m & c [illegible] varies X)

- It starts with the random value of m and c and that time value of m and c will change, relationship betn x & y will also change.
- Different m & c values it basically evaluate different possible lines before finding the best fit line.
- BFL could be that which goes through the maximum no. of data points and at same time it minimizes the distance betn other point of the line. & accordingly the error will be less.

(initially we can not get this BFL so algorithm executes multiple times.)



- So our algorithm initially creates multiple lines on the basis of different m and c values.

Memo No.
Date / /

- ~~from 1st all multiplied~~ but to find out this best fit line, and to have best value of m & c where we will be having least error.

- so there is a process or we can say can algorithm by which we can find out the best values of m & c which is called as "gradient decent algorithm".

L1 (Line) ML C1 >> suppose data point away from the line >> E1 (error) will be high

L2 ML C2 >> suppose data point bit away from the line >> E2 will be bit lower.

LBF MBF CBF >> Distance bet^n mat of data point are on line, data points above & below the line are at minimal distance

>> Ebf least

line Best fit 'm' best fit 'c' best fit

error Best fit.

Calculation of error (loss function / cost function)

- The loss is an error only. in our predicted value of m and c and our goal is to minimize the error to get most accurate value of m & c.

cost function measure how ML model performs. It calculating error between predicted values)

The cost fun^n of linear regression is mean squre error (MSE)

Residual error = in bet^n actual point & predicted point

Regression error = in bet^n mean value & predicted point

mean square error (mean square error)

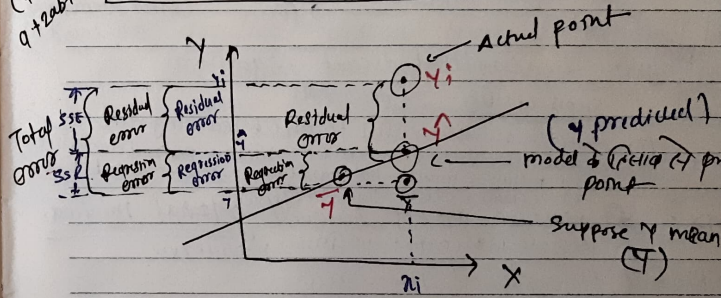
MSE mean square error $E = \frac{\sum (y_i - \bar{y}_i)^2}{n}$ (n = no. of element (or) total data point)

This is cost fun^n of linear regression

quadratic equation ⇒ $E = \frac{1}{n}\sum (y_i - 9)^2$

$E = \frac{1}{n}\sum (y_i - (mx_i + c))^2$

$(a+b)^2 \equiv a^2+2ab+b^2$



① $Y_{actual} - Y_{prediction}$ = Residual error (SSE)

② $Y_{prediction} - Y_{mean}$ = Regression error (SSR)

③ Total error = $Y_{actual} - Y_{mean}$ = Residual error + Regression error (SST) = SSE + SSR

SSE ⇒ Sum of square error ⇒ Residual ⇒ $y_i - \hat{y}$ ⇒ $\Sigma(y_i - \hat{y})^2$

SSR ⇒ Sum of sqrs ↓ Regression error ⇒ $\hat{y} - \bar{y}$ ⇒ $\Sigma(\hat{y} - \bar{y})^2$

SST ⇒ Sum of sqrs ↓ Total error / MSE ⇒ $y_i - \bar{y}$ ⇒ $\Sigma(y_i - \bar{y})^2$ ⇒ SSE + SSR

MSE ⇒ $\frac{SSE}{n}$ cost fun^n of linear regression

Gradient decent Algorithm :-

Convex shape

Bold doconward algorithm ... Partial derivative ... learning step ... Partial derivative है।

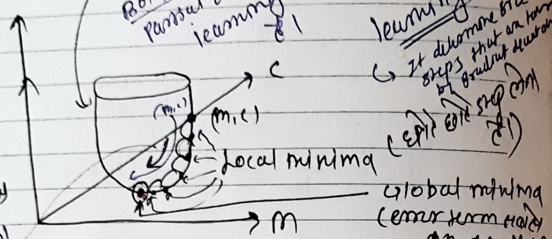


fig. Gradient decent Algorithm (doconward)

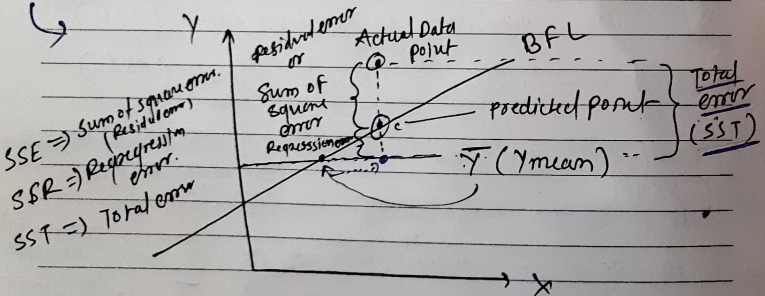
Steps of GDA

1. Initialise random value to m, c & learning rate
2. Create new value of m & c => (using partial derivative of cost fun^n w.r.t m & c)
   - 2.1 m(new) = m(old) − learning rate × p.d(m)
   - 2.2 c(new) = c(old) − learning rate × p.d(c)

- m & c का value goes toward global minima.

3. Repeat till cost fun^n (MSE) near to zero.

- To findout the global minima



- Total error is distance betn actual data point or y mean / expected value.

Day-59

Sum of Regression square => SSR
Sum of square error => SSE
Sum of square Total error => SST

Memo No.
Date 30/05/2022

Total error e.g.

~~people~~ people weight
one people weight
The difference betn them called total error.

- error means simply variance.
- The gradient decent work in finding the best fit line, where the BFL is that line where this error is minimum.

(1) So then model is most fit, it means all data points should lie on BFL, where SSE = 0 &
SSR should be equal to SST i.e. SSR/SST = 1

(2) If it is a poor fit / It mean SSE is large and SSR/SST will be close to zero.

- $\frac{SSR}{SST}$ it is nothing but R^2 or we can say R squared value.
- It is also called as Coefficient of determination.

Coefficient of determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

ideal value of $R^2 = 1$

R^2 lies betn 0 to 1

R lies betn −1 to +1

- SSE is always less than total error (SST) i.e SSE < SST

R^2 = 0 to 1
R = -1 to +1

Memo No.
Date / /

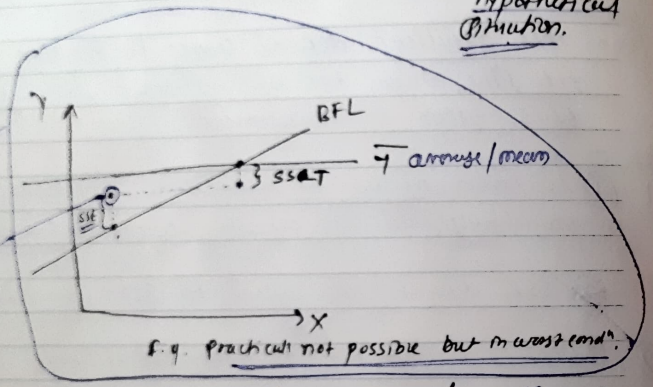
- SSE is always less than total error (SST)

↳ when SSE = 0, $R^2 = 1$ >> This is an ideal case.

+ when SSE < SST, R^2 = 0 to 1 >> This is an practical case.

Always

when SSE > SST, R^2 = negative value >> This is very hypothetical situation.



Y = output

Actual data point

- when BFL is worst than average/mean line, in that case SSE will be greater than SST.

- R^2 is used to check the goodness of BFL, For feature which good predictors, R^2 will be high. For feature which bad predictor, R^2 will be very low/close to zero.

Disadvantage of R^2

R^2 is not a good matric for evaluation of models, because it easily used to get impacted by Fluke relationships.

↳ Sometimes value of R^2 get increases for a bad predictor as well, in that case we go for an another concept called adjusted R-square value.

Adjusted R^2

Memo No
Date / /

	R^2	Adjusted R^2
Initially	0.85	0.84
Good predictor	0.87	0.86
Bad predictor	0.875	0.70

Some time value of R^2 increases for a bad predictor as well, in that case we go for an another concept called adjust R^2 value, which gets increase only when the good predictors has been to the model.

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$$

R^2 = R-squared = coefficient of Determination
n = no. of samples/rows in the dataset
p = no. of predictors/feature.

minimize the disadvantage of R^2

Adjusted R^2 get's increased only when the good predictor

R = coefficient of co-relation

Assumptions of Linear Regression

1. Assumptions of linearity >> There should be a linear relationship betn Dependent & independent Variable.
   - if R is approx. zero >> Non-linear. i.e. there is no linear relationship. It does not mean that there is no relationship at all.
   - It is a property of mathmatical relationship. which graphically represented as a straight line.

Linear relationship bet DV & IDV

plot scatter plot

2. Small/no multicolinearity
   - It is a concept in which there is high co-relation between two independent variable.
   - Due to multicolinearity it is difficult to get the relationship betn target variable & predictor variable. or independent variable.
   - x_1, x_2 >> independent variables (linearity in betn x_1 & x_2)
   - Y >> dependent variable

heatmap

I 0.3 to I 0.7

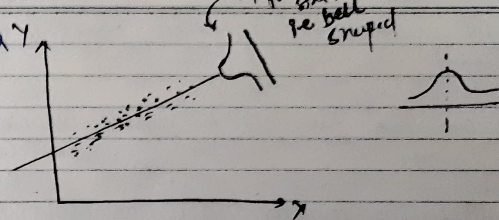
small or no multicolinearity in betn x_1, x_2 in IDV

3. Normal distribution in the error terms
   - There should be normal distribution in error terms.

Normal distribution of error & bell shaped curve, mean = 0, std = 1

Kele plot, Q-Q plot

analytical: 1. Shapiro 2. KS test test 3. Normal test



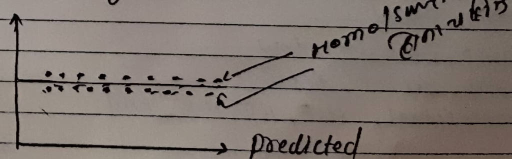
If we check variance on x-axis, we should get bell-shaped value (ideally)
means we should get as much good bell shaped curve as possible.
mean = 0, std = 1

4. Homoscadasticity: (Same Variance)

Homo >> Same
Scadasticity >> Variance (error)

- error terms should be as per the phenomenon of homoscadasticity.

same error term for all values at independent variable.



- error term should be almost same for all values of independent variable.

R = coefficient of correlation -1 to +1 R^2 = 0 to 1

Memo No.
Date / /

How to understand whether there is multicolinearity or not?

↳ There are few concepts by which we can get this thing:

1. Co-relation matrix (sns.heatmap(df.corr(), annot=True)) → R value
   # we have seen by using (corr())
2. VIF (range of VIF is 1 to ∞) >> any possible one value

if VIF ranges 1 to 5 >> independent to each other / moderately independent

if VIF ranges is greater than other is an issue of multicolinearity (or) we can say that independent variable are highly dependent.

VIF = Variance Inflation Factor

$VIF = \frac{1}{1-R^2}$

1 to 5 -- moderately correlated
>5 --- Highly correlated
1 --- Not correlated

if VIF = 1 to 5 >> completely independent

```
import pandas as pd
import numpy as np
```

Day-55

Ordinal value
1) high, medium, low
2) good, better, best
3) highschool, HSE, BE, ME, phd

Nominal value
pune, Delhi, mumbai etc.

Day-55

Encoding (we are discussing about categorical variable) (from sklearn.preprocessing import LabelEncoder)

label encoding
one hot encoding

i.e categorical variable convert into int/float value

Temperature (original value) → low, medium, high

Temperature (after applying label encoder) → 0, 1, 2

① Label encoding : - It is applied on target variable / y-axis (not on x-axis)

(original value)	(After applying label encoder)
high	2
medium	1
low	0

e.g	(original value)	(after applying label encoder)
	Setosa	0
	versicolor	1
	virginica	2

② One hot encoding

e.g	City_pune	City_Delhi	City_mumbai
pune	1	0	0
Delhi	0	1	0
mumbai	0	0	1

① Nominal encoding ↳ no need to worry about arrangement of category
② ordinal encoding ↳ Rank
We need to worry about rank. Because category will change on basis of rank

e.g - label encoding

```
df = pd.DataFrame({'city': ['pune', 'mumbai', 'Delhi', 'Delhi',
                            'pune', 'mumbai']})
df
```

	city
0	pune
1	mumbai
2	Delhi
3	Delhi
4	pune
5	mumbai

e.g - one hot encoding

```
df_city = pd.get_dummies(df['city'],
                         prefix='city')
df_city
```

	city_Delhi	city_mumbai	city_pune
0	0	0	1
1	0	1	0
2	1	0	0
3	1	0	0
4	0	0	1
5	0	1	0

Day-59

Memo No.
Date 6/6/22

this algorithm is the best suite for our model building.

8. Linear regression is prone to over-fitting, but with the help of some dimentionality reduction technique, Regularization methods (L1 and L2) it can avoid.

Disadvantages of Linear Regression Algorithm

1. It gets effect because of outliers.
2. It have many assumptions.
3. sometimes a lot of feature engineering is required.
4. Often get overfit.
5. It is sensitive to missing values.

Applications

1. House price prediction
2. temperature prediction
3. salary prediction
4. stock price prediction
5. weather forecasting
6. cricket score prediction.

Mo Tu We Th Fr Sa Su

used for only classification problem not for regression.

Day-60

classification problem means dependent variable in categorical form.

Linear regression is used when our dependent variable is continuous in nature for example weight, weight etc & in contrast.

Logistic regression is used when the dependent variable is binary or limited for example yes/no

2) Logistic Regression

- It is supervised type of machine learning algorithm
- It is used for the classification problem and not for regression problem. still the name given is logistic regression. (categorical form (D.V))

Q- why the name logistic regression & not logistic classifn

Ans - Because of at the back end it also uses linear model (or) we can also say that classification method uses the same concept of linear regression that is the reason it is called as logistic regression.

- The word logistic represent here, the log loss function / cost funn (they shows error)

- Target variable should be in the categorical form. It can be two or more categories.

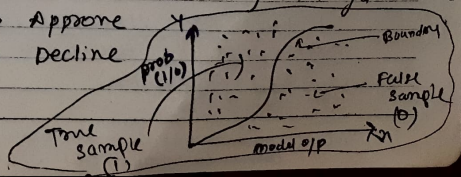
It does have 2 categories → 1) spam / not spam, 2) yes / no → ① binary class classifn

It does have more than two categories → 1) low / high / medium >> ordinal, 2) mango / apple / banana >> multinominal → ② multiclass classifn

① Binary class classification

e.g. Suppose we do have two classes in target variable

class1 >> Approve
class2 >> Decline



cost function (or) log function ... they shows the error

linear regression में => cost funⁿ (MSE) है।

logistic regression में => logs funⁿ (logloss) है।

(इसमें ~~...~~ logistic regression ... linear model ...)

[model.predict >> class]

It can also be used to find out probability with which an observation belong to either class 1 or class 2

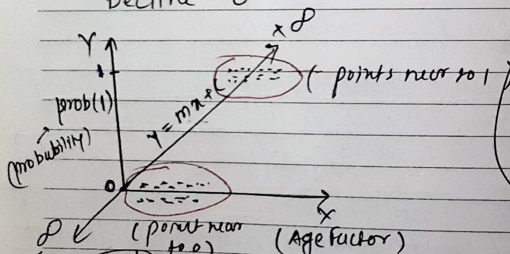
[model.predict_probe >> it predict the probability not class]

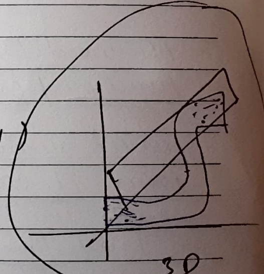
Decision tree, logistic regression, Naive bayes algorithm.

e.g Loan Application status

Approve - 1

Decline - 0





Linear model ~~में~~ probability ~~BFL line should be~~ always towards +∞ & −∞

- But here BFL line not came due to decision making tree. so we have solution that is sigmoid function.

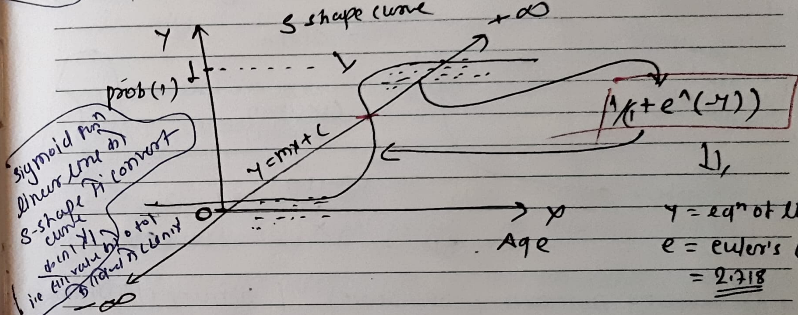
i.e in linear model ~~BFL~~ line cross the boundary of class ~~to not get proper~~ but BFL not came due to decision

$$\frac{1}{1+e^{\wedge}(mx+c)}$$

$$S(y) = \frac{1}{1+e^{-y}}$$

Sigmoid function equation

[~~1+e^{-y}~~ $\frac{1}{1+e^{-y}}$ => $1/(1+e^{\wedge}(-y))$]



y = eqⁿ of line

e = euler's constant = 2.718

- Sigmoid function simply tries to convert independent variable into an expression of probabilities that ranges in between 0 & 1 with respect to dependent variable.

- In linear model probability ~~varying~~ ~~going~~ always towards −∞ & +∞.

- Sigmoid function acts as an activation function in machine learning which is used to add non-linearity in a machine learning model. In simple words it decides which value to pass as o/p & what not to pass

Day - 61

Memo No.

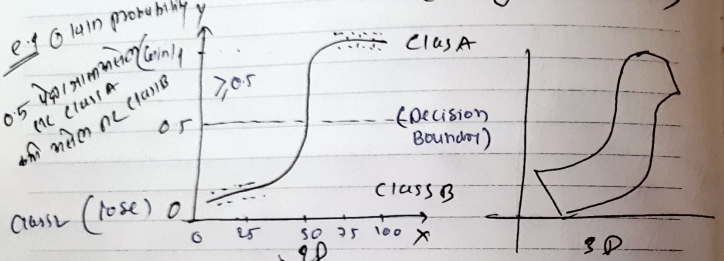
Date 09/06/22

S(Y) = 1 / (1 + e^-Y)

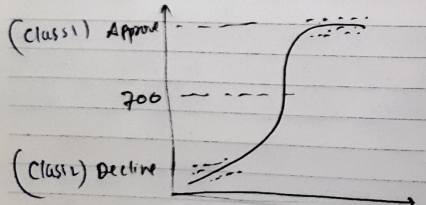
- With the help of sigmoid function it converts the liner line to S-shaped curve which is called as sigmoid curve.

(binary class classifn)

e.g. ① (with probability)



e.g. ② Banking Domain (CBil score) (binary classification)



If CBil score - >700 >> Approve
If CBil score - <700 >> Decline
even though CBil score >700 >> Decline (salary/income)

- Binary class classification either ~~objects~~ will belong to Class A or Class B

② Multiclass classification

multiclass classification How it works

C1
C2
C3

It uses a concept called OVR >> one verses rest. (one verses all)

C1 or Not in C1 (C1 Vs C2 and C3)

either C1 or not in C1
C2 or not in C2
i.e C1 Vs C2 & C3

working

- logistic regression is about two building steps

① The propensity to belong to class1 i.e P(Y/X)

② It basically uses a cut off or threshold to decide the class

③ It finds a DFL/plane, then it sends through sigmoid transformation and convert it into S-shaped curve.

- Shut the S-shaped which has been formed can be shifts to right-left and upward & downward

It happens on the basis of m and C value.

up and down >> on the basis of C value

right & left >> m value.

- So basically here with values of m & C we get to know that which sigmoid surface/curve/plane is best for our model.

$$-\frac{1}{M}\sum_{i=1}^{M} y_i \cdot \log(p(y_i)) + (1-y_i) \cdot \log(1-p(y_i))$$ =>

Memo No.
Date / /

And to actually find out which is the best sigmoid surface for it uses one function which is called logloss function.

Logloss function

- To find out which is the best sigmoid surface, it uses one function which is called as logloss function.

Logloss is indicative of how close the prediction probability is to the corresponding actual/true value.

$$\text{Logloss} = \frac{-1}{N}\sum \text{for all observation } (y_i (\log p_i) + (1-y_i)\log(1-p_i))$$
$$= -\frac{1}{N}\sum y_i \log(p_i) + (1-y_i)\log(1-p_i)$$

Y - Target variable, which ranges in both of 1
p - is our predicted probability.

$$y_i (\log p_i) + (1-y_i)\log(1-p_i) \Rightarrow y_i(\log p(y_i)) + (1-y_i)\log(1-p(y_i))$$

probability of 1 | probability of 0

Case 1: Y=1, P=1 (high)
↳ correct classification

p(yi) is the probability of 1
1-p(yi) is the probability of 0

Memo No.
Date / /

(Y=1, p is high >> second part our eqn will be 0.
yi(log pi)
and as here p is high then log pi = 0
The whole logloss function will be 0
This is ~~actually~~ basically a correct classification.

import numpy as np.

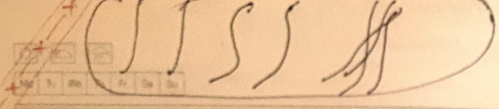
np.log(1)	np.log(0)
↳ 0.0	↳ -inf

Case 2: Y=0, P=0 (low) => correct classn

Y=0, first part will be 0, p is low
(1-yi) log(1-pi)
↳ log(1) = 0
This is basically a correct classification

Case 3: Y=1, P=0 => incorrect classn

Y=1, p is low
yi(log pi)
logloss function will not turn to 0, we will get some value here.
This is incorrect classification



← Sigmoid Curve

Memo No.
Date / /

Possibility

Case 4 ; $y=0$, $p=1$ ⇒ incorrect classification.

$y=0$, p is high

$(1-y_i)\log(1-p_i)$ ⇒ $1 \cdot \log(1-1)$

Logloss function will not turn to 0 here.
This is incorrect classification

Day-62

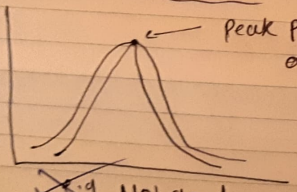
Assumptions of logistic Regression

1. Dependent variable should be catagorical in nature.
2. No multicolinearity. (in bet" independent variable)
3. It construct linear boundries.

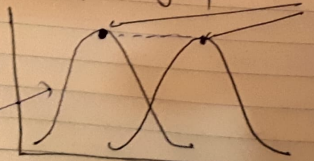
(It means

Peak and distributed curve of two variables should be seperated from each other

Peak point should not overlap.



e.g Not good



Distribution curve

Peak point should be seperate

e.g Very good

Day-62

Memo No.
Date 10/6/22

Peaks and distribution curve of two variables should be seperated from each other. If they are overlapped with each other, logistic regression will not do good.)

4. Data size should be large (No. of observations should be large)

VIMP in classification problem

Evaluation Metrics

VIMP

Performance Metric

↳ various evaluation metrics ⇒ Evaluation metrics quantify the performance of a predictive model (or) measure the quality of ML model.

1. Accuracy
2. Confusion matrix
3. Log-loss ⇒ $-\frac{1}{N}\sum y_i \cdot \log p(y_i) + (1-y_i) \cdot \log(1-p(y_i))$
4. precision and Recall
5. F1-Score ⇒ Fb(bita) ⇒ bi bita( )
6. AUC-ROC curve

1. Accuracy ⇒ . Ratio bet" no. of correct predictions & total no. of predictions

- Accuracy is measure of how often the classifier makes the correct prediction.
- It is a ratio between the no. of correct predictions and the total no. of predictions.

Accuracy Score = $\frac{TP+TN}{TP+FP+TN+FN}$