Dr. Devender sir

Dr. Devender sir

Supervised Learning / ML

(1)

linear Regression: It is a statistical model that attempts to show the relationship between two variables, with the linear equation.

linear Regression algorithm (LRA) using Least Square method (LSM) check its goodness of fit or how close the data is to the fitted regression line using the R^2 method. And then finally optimize it using the Gradient Descent method.

— What is Regression?

"Regression Analysis is a form of predictive modelling Technique which investigates the Relationship between a dependent and Independent Variable".

It involves graphing a line over a set of data points that most closely fits the overall shape of the data or Regression shows the changes in a dependent Variable on the y-axis to the changes in the explanatory Variable on the x-axis.

— **The uses of Regression:** There are three major uses.

(1) Determining the strength of predictors
(2) Forecasting an effect
(3) Trend forecasting

Regression might be used to identify the strength of the effect, that the independent variables have on the dependent Variable.

Eg: the strength of Relationship between Sales and marketing spending. or
the relationship between age and income

(2) Forecasting an effect: Here the Regression used to forecast effects or impact of changes. Here Regression Analysis helped us to understand how much the dependent variable changes with the change in one or more independent variable.

Eg: how much additional sale income will I get for each thousand dollars ~~spend~~ spent on marketing.

(3) Trend forecasting: In this the RA to predict trends and future values. The RA can be used to get point estimates.

Eg: What will be the price of Bitcoin in next six months?

~~Diff~~ Linear Regression Vs Logistic Regression

Basis	LR	Logistic Regression
(1) Core concept	The data is modelled using a straight line	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
(2) Used with	Continuous variables	(2) Categorical variable
(3) O/P/prediction	Value of the variable	(3) Probability of occurrence of event
(4) Accuracy and goodness of fit	Measured by loss, R squared, Adjusted R squared etc.	(8) Accuracy prediction, Recall, F1 score, ROC curve, confussion Matrix (CM), etc

Scanned with OKEN Scanner

LR Selection Criteria:

(1) Classification and Regression capabilities:

Regression models predict a continuous variable such as sales made on a day or predict the Temperature of a city their reliance on a polynomial like a straight line to fit a data set posses a real challenge when it comes towards building a classification capability.

$\boxed{\text{LR is not good for classification model}}$

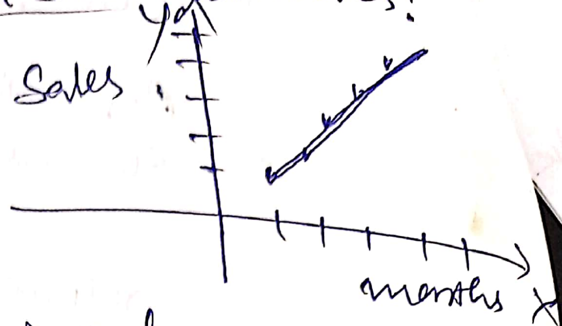
(2) Data Quality: Each missing value removes one data point that could optimize the regression. in SLR the outliers can ~~specifically~~ significantly distrupt the outcome

(3) Computational complexity: LR is not computationally expensive as compared to the Decision Tree or the clustering algorithms. The order of complexity for 'n' training examples and 'x' features is $O(x^2)$ or $O(x n)$

(4): Comprehensible and Transparent: The LR are easily comprehensible and Transparent in nature. they can be represented by a simple mathematical notation to any one and can be understood very easily

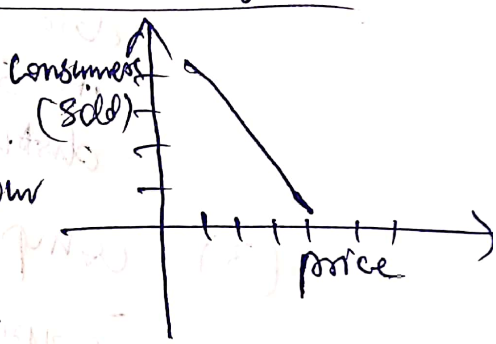
Where is LR used?

(1) Evaluating Trends and Sales estimates:



If a company sales have increased steadily every month then conducting a linear analysis on the sales data with monthly sales on y-axis and time on the x-axis.
It predicts the upward Trends in the sales after creating the Trendline the company use the slope of the lines to focus sales in future months.

(2) Analyzing the Impact of price changes:



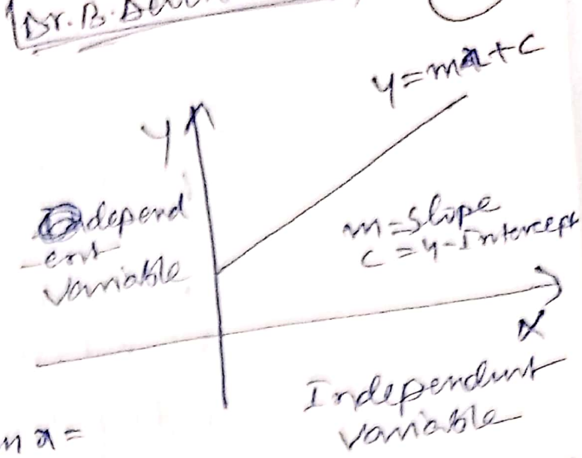
The impact of price changes with LR can be used to analyze the effect of pricing and consumer behaviour.

If company increases the price of a certain product how the consumers using is imp.
- If more increases less consuming.

This will help us future pricing decisions

(3) Assessment of risk in financial services and Insurance Domain: LR can be used to analyze the risk.
Eg: Health Insurance company by plotting the no. of claims per customer against his age
Old customers can claims more insurance claim.

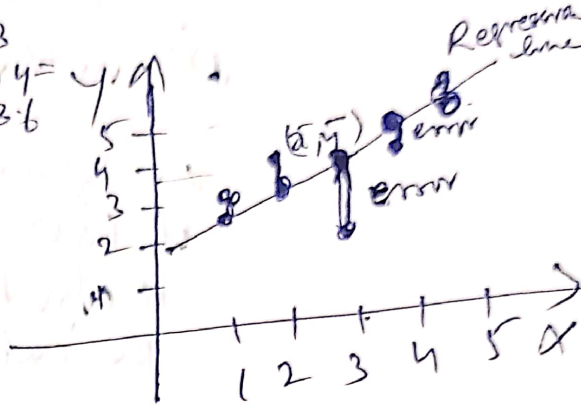
Understanding LRA:



Ex:

x	y
1	3
2	4
3	2
4	4
5	5
$\bar{x}=3$	$\bar{y}=3.6$

Mean x =
$\bar{x}=3$
mean y=
$\bar{y}=3.6$



We have to find the Regression line.

$$y=mx+c \quad \text{—(1)}$$

$$m=\frac{\Sigma(\bar{x}-x)(y-\bar{y})}{\Sigma(x-\bar{x})^2}=\frac{4}{10}=0.4.$$

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	3	-2	-0.6	4	1.2
2	4	-1	0.4	1	-0.4
3	2	0	-1.6	0	0
4	4	1	0.4	1	0.4
5	5	2	1.4	4	2.8
$\bar{x}=3$	$\bar{y}=3.6$			$\Sigma=10$	$\Sigma=4$

Apply $m, \bar{y}, \bar{x}$ values in the eqn (1) then $c=2.4$.

Regression line $\doteq y=0.4x+2.4$

Now predicted values for y for $x=\{1,2,3,4,5\}$

$y=0.4\times1+2.4=2.8$
$y=0.4\times2+2.4=3.2$
$y=0.4\times3+2.4=3.6$
$y=0.4\times4+2.4=4.0$
$y=0.4\times5+2.4=4.4$

the predicted line is formed for (1, 2.8), (2, 3.2), (3, 3.6), (4, 4), (5, 4.4)

Now calculate the Distance between Actual and predicted Value and aim is to reduce the distance.

The line with the least error will be the line of LR line, and It will also be the best fit line

finding the Best fit line:

R^2 method (R-square):

R-squared ~~method~~ Value is a statistical Measure of how close the data are to the fitted regression line it also known as "Coefficient of Determination" or the "coefficient of multiple determination".

Here distance = Actual − mean

Vs

distance predicted − mean

$$\boxed{R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2}}$$

x	y	$y-\bar{y}$	$(y-\bar{y})^2$	y_p	$y_p-\bar{y}$	$(y_p-\bar{y})^2$
1	3	−0.6	0.36	2.8	−0.8	0.64
2	4	0.4	0.16	3.2	−0.4	0.16
3	2	−1.6	2.56	3.6	0	0
4	4	0.4	0.16	4.0	0.4	0.16
5	5	1.4	1.96	4.4	0.8	0.64

$\bar{y} = 3.6$ $\Sigma = 5.2$ $\Sigma = 1.6$

$$R^2 = \frac{1.6}{5.2} = 0.3$$

if R^2 value ≈ 1 then it is a good fit line

Here the prediction line is not good fit

Lower R^2 values are not good fit lines.

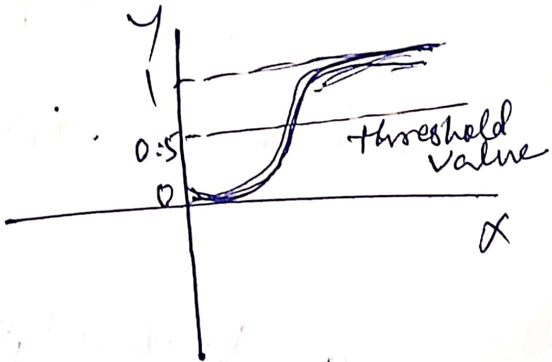
Logistic Regression:

Logistic Regression produces results in a binary format which is used to predict the outcome of a categorical dependent variable. So the outcome should be discrete/categorical such as 0 or 1
yes or No
True or false
High or low

Logistic Regression Curve:

Sigmoid curve is used.

It converts any value from $-\infty$ to $+\infty$ to a discrete value which is logistic value.



The values < threshold value it is considered as '0' otherwise it is considered as '1'.

Threshold values indicate the probability of winning or loosing.
winning = 1; loosing = 0.

The logistic Regression equation:

It is derived from straight line equation.

$y = c + B_1x_1 + B_2x_2 + \cdots$ range from $-\infty$ to $+\infty$
(more than one variables)

logistic Regression equation:

$y = c + B_1x_1 + B_2x_2 + \cdots$ y can be only '0' to '1'.

$$\frac{y}{1-y} \begin{cases} y=0 \text{ then } 0 \\ y=1 \text{ then } \infty \end{cases}$$

Now the range of 'y' is 0 to ∞.

$$\log\left(\frac{y}{1-y}\right) \Rightarrow y = c + \beta_1 x_1 + \beta_2 x_2 + \cdots$$

Logistic Regression Equation.

Here y range is $-\infty$ to $+\infty$

Differences:

Linear	Logistic
1. Continuous Variables	1. Categorical variables
2. Solves regression problems	2. Solves classification problems
3. straight-line graph	3. S-curve (Sigmoid)

→ Some Use-Cases:
1. weather prediction — linear Regression as well logistic Regression
2. Classification problems

1. Vertebrates → birds, fish, mammals, Reptiles } It can perform multi-class classification

2. Determines the illness: If a patient goes for checkup to Hospital, Doctor will do various tests and comes to conclusion with predictions. (Sugar levels, BP, Age, previous medical History etc)

→ Implementing Logistic Regression:

1. Collect the data → 2. Analyze the data → 3. Clean the data → 4. Train & Test data → 5. Accuracy check

Logistic Regression in Python | Logistic Regression Example | Machine Learning...

Titanic Dataset

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5 2151	8.0500	NaN	S
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S

edureka! 14:35 / 53:40

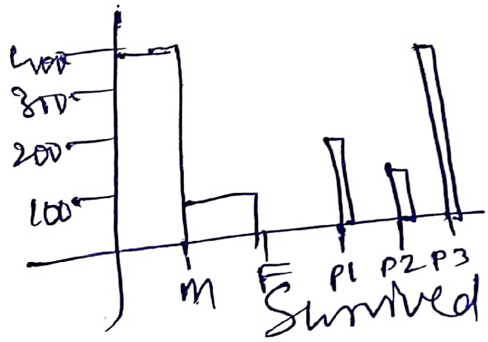
Data Science Certification Training

USE-CASE: Titanic ship data Analysis:

Explore titanic Dataset and explore about the people, both those who survived and those who did not.
What factors made people more likely to survive the sinking of the Titanic?

No. of passengers = 891

Analyze the data:



Unsupervised Learning

Clustering: clustering is the process of dividing the datasets into groups, consisting of similar data points.

- points in the same group are as similar as possible
- points in ~~the~~ different group are as dissimilar as possible.

Ex:(1) group of dinners in a restaurent
Ex: (2) Items arranged in a mall
Ex:(3) Amazon – Recommendation system
Ex:(4) Netflix – " movies

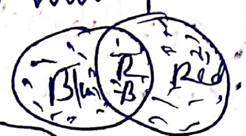
Types of Clustering: (1) Exclussive clustering
(2) overlapping "
(3) Hierarchical "

(1) **Exclusive Clustering:** (1) Hard clustering
(2) Data point / item belongs to one cluster exclusively

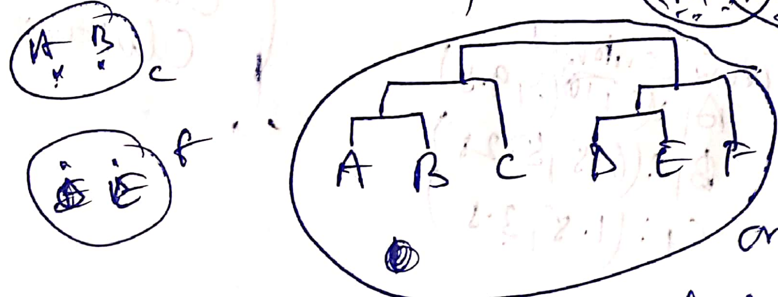
Ex: K-Means Clustering

(2) **Overlapping Clustering:** (1) Soft Cluster
(2) Data point / item belong to multiple cluster

Ex: fuzzy / c-means clustering



(3)



one cluster

Hierarchical clustering

K-Means clustering:

K-means is a clustering algorithm whose main goal is to group similar elements or data points into a cluster.

Note: 'K' in K-means represents the no. of clusters.

Eg: Suppose that the data mining table is to cluster points into three clusters.

Dataset: $A_1 = (2,10)$; $A_2 = (2,5)$, $A_3 = (8,4)$
$B_1 = (5,8)$; $B_2 = (7,5)$; $B_3 = (6,4)$
$C_1 = (1,2)$; $C_2 = (4,9)$

Initially A_1, B_1, C_1 points are centers of each cluster. Use Euclidian distance formula. Find distances from each centroid.

Initial centroids:
$A_1 : (2,10)$
$B_1 : (5,8)$
$C_1 : (1,2)$

Data points			Distance to 2, 10	Distance to 5, 8	Distance to 1, 2	Cluster I_1	New cluster I_2	I_3
A_1	2	10	0.00	3.61	8.06	1	1	1
A_2	2	5	5	4.24	3.16	3	3	3
A_3	8	4	8.49	5.00	7.28	2	2	2
B_1	5	8	3.61	0.00	7.21	2	2	1
B_2	7	5	7.07	3.61	6.71	2	2	2
B_3	6	4	7.21	4.12	5.39	2	2	2
C_1	1	2	8.06	7.21	0.00	3	3	3
C_2	4	9	2.24	1.41	7.62	2	→ 1	1

final cluster

New centroids are
$A_1 : (2,10)$
$B_1 : (6,6)$
$C_1 = (1.5, 3.5)$

Current centroids:

New centroids:
$A_1 : (3, 9.5)$
$B_1 : (6.5, 5.25)$
$C_1 : (1.5, 3.5)$

Cluster 1: A_1, B_1, C_2
Cluster 2: A_3, B_2, B_3
Cluster 3: A_2, C_1

Reinforcement Learning:

Bellman Equation:

The Key elements used in:

- ~~Action~~ performed by the agent is 'a'
- State occurred by performing the action is 's'
- The Reward/feedback obtained for each good and bad action is 'R'
- The discount factor is 'γ' (Gamma)

then

Bellman equation $= V(s) = \max\left[R(s,a) + \gamma V(s')\right]$

Where: $V(s)$ = Value calculated at a particular point

$R(s,a)$ = Reward at a particular state 's' by performing an action 'a'

γ = Discount factor

$V(s')$ = The value at the previous state.

Here we are taking the max of the complete values ∵ agent tries to find the optimal value

$\gamma = 0.9$ assume

find the values at each state start from S_3.

for S_3 block

$V(s) = \max(R(s,a) + \gamma V(s'))$

$= \max(1 + 0.9(0)) = 1$

$S_2 = \max(0 + 0.9 \times 1) = 0.9$

$S_1 = \max(0 + 0.9 \times 0.9) = 0.81$

$S_5 = \max(0 + 0.9 \times 0.81) = 0.73$

0.81 S_1	0.9 S_2	1 S_3	Goal	R=1
0.73 S_5	Locked S_6	0.9 S_7	Danger	R=-1
0.66 S_9	~~0.73~~ 0.59 S_{10}	0.81 ~~0.53~~ S_{11}	~~0.73~~ 0.68 S_{12}	

finally whenever Robot at any state it can find optimal path.

Q-Learning

Policy is undertaken by agent
(agent observes the given state and selects best possible action)

Policy: $\phi^{\pi}(s_t, a_t)$ where s_t - state at time 't'
a_t - action at time 't'

Reward: Scalar quantity
correct action → reward

Penalty: commonly called as -ve reward.

Q-learning mainly depends on two factors:

1. Q-function:

$$Q^{\pi}(s_t, a_t) = E\left[R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{n} R_{t+n} \mid s_t, a_t\right]$$

⇓

Bellman Equation.

2. Q-Table: combination of actions and states.

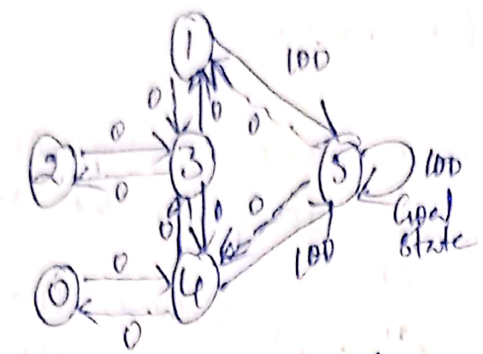
Ex: In a game ⟺
actions: up, down, left, right
state: start, end, reward, health, etc

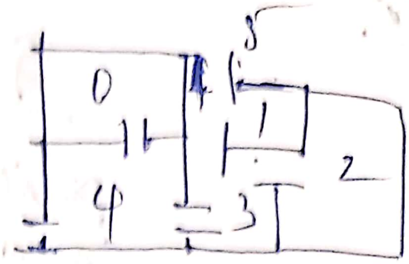
Steps followed:

(1) Exploration: explore all possible paths
(2) Exploitation: best possible path is identified
(3) Initialize a Q-table = 0
(4) Choose Action
(5) perform action
(6) measure reward
(7) update a Q-table

Q-learning

Ex:





We can put the state diagram and the instant reward values into the following reward table "matrix R"
the -1's represents null values.

Action

States	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

learning rate $= \gamma = 0.8$

Q =

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

$\Rightarrow$

	0	1	2	3	4	5
0	0	0	0	0	80	0
1	0	0	0	64	0	100
2	0	0	0	64	0	0
3	0	80	51	0	80	0
4	64	0	0	64	0	100
5	0	80	0	0	80	100

$Q(\text{state}, \text{action}) = R(\text{state}, \text{action}) + \gamma * \text{Max}[Q(\text{next state}, \text{all actions})]$

$Q(1,5) = R(1,5) + 0.8 * \max(Q(5,1), Q(5,4), Q(5,5))$

$= 100 + 0.8 * 0 = 100$

$Q(3,1) = R(3,1) + 0.8 * \max(Q(1,3), Q(1,5)) = 0 + 0.8 \times 100 = 80$

⋮

like tracing the best sequence