→ Probability
→ Bayse Theorem

L-1

More Data, More Questions, Better Answer

More data : features

Data ~~Ae~~ should be Authenticated.

- Machine Learning
  - → Unsupervised Learning → Clustering
  - → Supervised Learning → Classification
    - → Regression

→ Generalisation

Designing a Learning System

→ Choosing the training experience.
  • direct and indirect feedback.
  eg : chess game

→ One key attribute is : Whether the training experience provides direct or indirect feedback regarding the choices made by the performance system.

→ A second important attribute of the training experience, is the degree to which learner controls the sequence of training example.
  • control you have over your data

Date
Page No.

→ The third important attribute of the training experience is: How well it represents the distribution of examples over which the final system performance must be measured.

→ Choosing the target function. (Representation of target function)

→ IV: Choosing a function approximation algorithm.

→ Estimate training value or final design

What is Supervised Learning?

i. The aim of supervised M/C learning is to build a model that makes predictions based on evidence in the presence of uncertainity.

ii. A supervised learning algorithm, takes a known set of input data and known responses to the data, which is the output:

And trains a model, to generate reasonable predictions for the response to new data.

~~What is~~

→ Classification
→ Regression

Classification:

This technique predicts direct responses.
eg: Email: Spam or Not Spam.
Medical field:
Bot Prediction

Regression

Continious predictions:
eg: Power Demand fluctuations.
Change in temperature.

Page No.

UNSUPERVISED Learning

→ There are no outputs.

→ Shows you the hidden patterns / structures for analysis

L-2

Simple Linear Regression

When we have a single input attribute and we want to predict a value, we will used Linear Reg.

- Works on eqn of line.

$$y = \underbrace{m}_{\text{slope}}x + \underbrace{c}_{y\text{ intercept}}$$

$$y = \underbrace{B_0}_{\text{Intercept}} + \underbrace{B_1}_{\text{Slope}} * x \quad \therefore \text{Coeff.}$$

· William Stalling : Crypto

eg:

1	x	y
2	1	1
3	2	3
4	3	3
5	4	2
6	5	5

The goal of this algorithm is to find the best estimates for the co-efficients to minimize the errors in predicting Y from X.

→ First we estimate the value of B_1:

$$B_1 = \frac{\text{Sum}\left((x_i - \text{mean}(x)) * (y_i - \text{mean}(y)) \right)}{\text{Sum}\left((x_i - \text{mean}(x))^2 \right)}$$

$$\boxed{B_0 = \text{mean}(y) - B_1 * \text{mean}(x)}$$

$\frac{1}{2}$ (P

Date — Charlie
Page No.

Estimating the slope (B_1)

$$mean(x) = \frac{1}{n} * sum(x)$$

$$= \frac{1}{5} * 15$$

$$= 3$$

$$Mean(y) = \frac{1}{5} * 14$$

$$= 2.8$$

1	x	~~x~~ mean(x)	x - mean(x)
2	1	3	-2
3	2	3	-1
4	3	3	0
5	4	3	1
6	5	3	2

1	x	mean(y)	y - mean(y)
2	1	2.8	-1.8
3	3	2.8	0.2
4	3	2.8	0.2
5	2	2.8	-0.8
6	5	2.8	2.2

x-mean(x)	y-mean(y)	multiplication
-2	-1.8	3.6
-1	0.2	-0.2
0	0.2	0
1	-0.8	-0.8
2	2.2	4.4
		7.0

$(x-mean(x))^2$
4
1
0
1
4
10

$$B_1 = \frac{7}{10}$$

$$= 0.7$$

$$B_0 = 2.8 - \frac{7}{10}(3)$$

$$= 2.8 - \frac{21}{10}$$

$$= 2.8 - 2.1$$

$$= 0.7$$

$y = 0.7 + 0.7(x)$

$\Rightarrow y = 0.7 + 0.7(1)$

$= 0.7 + 0.7$

$= 1.4$ [1]

$y = 0.7 + 0.7(4)$

$= 0.7 + 2.8$

$= 3.5$ [2]

$y = 0.7 + 0.7(5)$

$= 4.2$ [5]

$y = 0.7 + 0.7(2)$

$= 0.7 + 1.4$

$= 2.1$ [3]

$y = 0.7 + 0.7(3)$

$= 2.8$ [3]

.75

Date Charlie
Page No.

RMSE =

sqrt (Sum $(p_i - y_i)^2$ / n)

p_i	predicted value y_i	actual value	$p_i - y_i$	
1.4	1		0.4	.16
2.1	3		−0.9	.81
2.8	3		−0.2	.04
3.5	2		1.5	2.25
4.2	5		−0.8	0.64
				3.90

$\frac{3.9}{5}$

= 0.78

$\boxed{= 0.88}$

Shortcut $B_1 = \boxed{\dfrac{\text{Corr}(X, Y) * \text{Stdev}(Y)}{\text{Stdev}(X)}}$

1. Range = Largest observation − Smallest observation.

2. Mean Deviation.

a) Mean Deviation for Discrete Frequency Distribution.

$$MD(\bar{x}) = \frac{\Sigma f_i (x_i - \bar{x})}{\Sigma f_i}$$

Mean Deviation for Grouped Data

$$M.D(\bar{x}) = \frac{\Sigma f_i |x_i - \bar{x}|}{N}$$

{ Then x_i will be midpoint of a class
eg (15−20) = 17.5 }

Mean Deviation for Ungrouped data:

$$M.D(\bar{x}) = \frac{\Sigma |x_i - \bar{x}|}{n}$$

Variance

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

n = no. of observations.

Standard Deviation

$$\sigma = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\text{variance}}$$

Standard Deviation for a discrete frequency

$$\sigma = \sqrt{\frac{1}{N}\Sigma f_i (x_i - \bar{x})^2}$$

Standard Deviation for a Grouped Data

$$\sigma = \sqrt{\frac{1}{N}\Sigma f_i (x_i - \bar{x})^2}$$

$\hookrightarrow$ midpoint

Standard Deviation

$$\sigma_x = \frac{h}{N}\sqrt{N\Sigma f_i y_i^2 - (\Sigma f_i y_i)^2} \qquad y_i = \frac{x_i - A}{h}$$

h = width of class Interval

Date
Page No.

Assumed mean

1 10-20
2 20-30
3 .
4 .
5 .

$\frac{5+1}{2}$ = 3rd no. observation odd

$\left\{ \frac{n}{2} \text{ or } \frac{1+last}{2} \right\}$ even

Coefficient of Variation

$= \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$

Scanned by CamScanner

Page No.

Q

Size (x_i)	1	3	5	7	9	11	13	15
frequency (f_i)	3	3	4	14	7	4	3	4

'Mean

→ find Deviation about Mean'

→ find Mean

$$M.D = \frac{\sum f_i |x_i - M|}{n}$$

$\sum f_i = 42 = n$

$\cancel{\{Mean = 8\}}$ $Mean = \frac{64}{8} = 8$

$$M.D = \frac{3(1-8) + 3(3-8) + 4(5-8) + 14(7-8) + 7(9-8) + 4(11-8) + 3(13-8) + 4(15-8)}{8}$$

$$= \frac{+21 + 15 + 12 + 14 + 7 + 12 + 15 + 28}{\cancel{8}\,42}$$

=

21
15
12
14
07
12
15
28
124
42

D-95

Q. Find the variance & standard deviation for the following data

57, 64, 43, 67, 49, 59, 44, 47, 61, 59

Mean = 55

$\sigma^2 = \frac{1}{10}[4+81+144+144+36+16+121+64+36+16]$

$= \frac{662}{10}$

66.2

$\sigma = 8.1$

Q.

Class-Interval	Frequency
4-8	3
8-12	6
12-16	4
16-20	7
	20

$\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}$

Mean $= \frac{(3\times 6) + (10\times 6) + (14\times 4) + (18\times 7)}{20} = \bar{x}$

$= 13$

Date

Page No.

~~$(3-13)^2 + (6-13)^2 + (6-$~~

x_i are midpoints

$$\frac{(6-13)^2 + (10-13)^2 + (12-13)^2 + (18-13)}{20}$$

$$= \frac{\cancel{3(25)+6(25)+4(6)+7(1)}}{20}$$

$$= \frac{\cancel{84}}{\cancel{20}} = \frac{147+150+36+7}{20}$$

$$= 19$$

55
435
450
140
225
159

1464

70

6

Charlie
Date
Page No.

Q. A Set of n values x_1 upto x_n has std dev = 6

the std. dev of n values $x_1 + K$, $x_2 + K$ unto $x_n + K$ .. will be?

Ans: 6

Q

Classes	Frequency
1-10	11
10-20	29
20-30	18
30-40	4
40-50	5
50-60	3

Mean?

Variation?

Standard Deviation.

$$\frac{(11 \times 5) + (29 \times 15) + (18 \times 25) + (4 \times 35) + (5 \times 45) + (3 \times 53)}{70}$$

$$= \frac{55 + 435 + 450 + 140 + 225 + 159}{70}$$

Scanned by CamScanner

*roper bta

Independent X_1	Independent X_2	Dependent Y
Highest Year of School	Motivation Measured	Annual Sales
12	32	350,000
14	35	399,675
15	45	429,000
16	50	435,000
18	65	433,000

	Mean	SD
X_1	15	2.236
X_2	45.5	13.164
Y	409,353	36,116.653

Correlation between X_1 & X_2 =

Correlation between X_1 & Y =

Correlation between X_2 & Y =

$$S_{xy} = \frac{S_{xy}}{\underbrace{S_x \, S_y}_{SD}} \rightarrow \text{Covariance}$$

$$\rightarrow \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

• Linear Regression

Date

Page No.

$$R_{xy} = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Correlation x_1 & x_2

~~(12-15)(32-45.5)+(14-15)(35-3~~

x_i	$x_i - \bar{x}$	x_2	$x_2 - \bar{x}$
12	12-15=-3	32	32-45.4=-13.4
14	14-15=-1	35	35-45.4=10.4
15	15-15=0	45	45-45.4=-0.4
16	16-15=1	50	50-45.4=4.6
18	18-15=3	65	65-45.4=19.6

$$\frac{(-3)(-13.4) + (-1 \times -10.4) + (4.6)(3 \times 19.6)}{5}$$

$$= \frac{40.2 + 10.4 + 4.6 + 58.8}{5-1}$$

~~= 22.8~~

$$\frac{22.8}{2.236 \times 13.164}$$

Corr $(x_1, x_2) = 0.968$

Corr $(x_1, \cancel{y}) = 0.880$

Corr (x_2, y) 0.772

- 9360

Strong rel between 3 variables

Date

Page No.

$$R = \sqrt{\frac{[(r_{y,x_1})^2 + (r_{y,x_2})^2] - [2 r_{y,x_1} r_{y,x_2} r_{x_1,x_2}]}{1 - (r_{x_1,x_2})^2}}$$

$= 0.774 + 0.595984 - [2\alpha \_ \_ \_]$

- p value { let 0.5 }
  - backward & forward elimination
- Adjusted R^2 and R.

LR - GD

Charlie

Date

Page No.

$y = ax + b$ LR

$y = a + bx_1 + b_2 x_2 \rightarrow$ MR

AS | Highest year of school | Motivation.

$$b_1 = \left[\frac{r_{y,x_1} - r_{y,x_2}\, r_{x_1,x_2}}{1-(r_{x_1,x_2})^2}\right]\left[\frac{SD_y}{SD_{x_1}}\right]$$

$$b_2 = \left[\frac{r_{y,x_2} - r_{y,x_1}\, r_{x_1 x_2}}{1-(r_{x_1,x_2})^2}\right]\left[\frac{SD_y}{SD_{x_2}}\right]$$

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

$$b_1 = \left[\frac{0.88 - (0.772 \times 0.968)}{1-(0.968)^2}\right]\left[\frac{36118.693}{2.236}\right]$$

$$= \left[\frac{0.132704}{0.062976}\right]\left[16152.36717\right]$$

$$= 2.107215 \times 16152.36717$$

$$= 34035$$

· RMS
· mean square
→ R → R^2

Date
Page No.

Most Useful Metrics

1) RMSE (Root Mean Square Error)

$RMSE = (MSE)^{1/2}$

It represents the sample standard deviation of the differences between predicted values and observed values.

2 → Mean Absolute Error

MAE is the average of the absolute difference between the predicted value, and observed value.
~~It is linear value.~~

Case 1: Actual Values = [2, 4, 6, 8]
& Predicted Values = [4, 6, 8, 10]

Case 2: Actual Values = [2, 4, 6, 8]
Predicted values = [4, 6, 8, 12]

MAE for case 1 = 2.0
RMSE for case 1 = 2.0

MAE for case 2 = 2.5
RMSE for case 2 = 2.5

R^2 [0.9854]
adj R^2 [0.9805]

Charlie
Date
Page No.

R^2 and Adjusted R^2

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n[\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-K-1}\right]$$

n : no. of points in your sample.
K : no. of independent regressors, no. of variables in your model excluding the constant.

X_1	Y (x)	(ascl : y) (y)	(xy)	x^2	y^2
1	2.1	2.3	4.83	4.41	5.29
2	4	4.08	16.32	16	16.6464
3	6.2	5.86	36.332	38.44	34.3396
4	8	7.64	61.12	64	58.3696
5	9	9.42	84.78	81	88.7364
				Σ 203.85	203.382

R^2 =
adj R^2 =

$\Sigma x = 29.3$ $\quad \Sigma y = 29.3$ $\quad \Sigma xy = 203.382$
$(\Sigma x)^2 = 858.49$ $\quad (\Sigma y)^2 = 858.49$

$$= \frac{5(203.382) - (29.3 \times 29.3)}{\sqrt{5[203.85 - 858.49][5(203.382)][858.49]}}$$

P values
for every predictors

Date
Page No.

x

$\sqrt{(1019.25 - 858.75)^2}$

$\frac{158.42}{160.8}$

$= 0.98519$

Log 2

2.206

4.22

5.766

7.78

9.32

$R^2 = 0.987$

$adj R^2 = 0.974$

$1 - \left[\frac{(1 - 0.9854)(5-1)}{5-1-1}\right]$

$= 1 - \left[\frac{0.0584}{3}\right]$

$= 1 - 0.019$

$= 0.9805$

- optimisation algo

Gradient Descent Algo

* It is an optimisation algo.
* In machine learning, we perform optimisation, on the training data and check its performance on a new validation data.

→ Squared Sum error

House Size sqrt(x)	1400	1600	1700	1875	1100	1550
House price $(Y)	245000	212000	279000	308000	199000	219000

2350	2450	1425	1700
405,000	324,000	319,000	255,000

• make curve

→ Linear Model

$Y_{pred} = a + bx$

Sum of Squared Error (SSE) =

$\frac{1}{2}$ Sum(Actual House Price − Predicted House Price)^2

Charlie Date Page No.

Step-1

Initialise the weights a and b with random values and calculate error SSE.

Step 2 Calculate the gradient that is change in SSE, when the weights ~~a~~ and b are changed by a very small value from their original randomly initialised value.

Step 3 Adjust the weights with the gradients to reach the optimal values, where SSE is minimised, ~~to~~

Step 4 Use the new weights for prediction and to calculate the new SSE

Step 5 Repeat steps 2 and 3 till further adjustments to weights, does not significantly reduce the error.

Charlie

Date

Page No.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \frac{1100-1100}{2450-1100} = 0$$

Min - Max Standardization

X	Y
0	0
0.22	0.22
0.24	0.58
0.33	0.20
0.37	0.55
0.44	0.39
0.45	0.54
0.57	0.53
0.93	1.00
1.00	0.61

Housing Data

~~1100~~	House Size (x)	Price (x)
1~~400~~	1100	199000
1~~600~~	1400	245000
	1425	319000
	1550	240000
	1600	312000
	1700	255000
	1700	279000
	1875	308000
	2350	405000
	2450	324000

0.1560125

Let

$a = 0.45$
$b = 0.75$

a	b	X	Y	$YP = a+bX$	SSE $= \frac{1}{2}(Y-Y_P)^2$
0.45	0.75	0	0	0.45	0.10125
		0.22	0.22	0.615	0.0780125
		0.27	0.58	0.63	0.001
		0.33	0.20	0.6975	0.125
		0.37	0.55	0.7275	
		0.44	0.39	0.78	
		0.45	0.34	0.78	
		0.57	0.53	0.8775	
		0.93	1.00	1.475	
		1.00	0.61	1.2	
					Σ 0.677

$$\partial SSE / \partial a = -(Y - YP)$$

$$\partial SSE / \partial b = -(Y - YP)X$$

$$a = a - r * \underbrace{\partial SSE / \partial a}$$

r = learning rate = 0.01

$$b = b - r * \partial SSE / \partial b$$

$a = 0.12$

$b = 0.73$

$\vdots$

$\epsilon = 0.553$

Logistic Regression

Data set should be binary.

dependent variable must be 'Binary'.

Attendance	Mid term	Pass/fail	%
		P	
		F	

→ type of classification.

→ solved as regression.

S.No	X1	X2	Y
1	2.781	2.550	0
2	1.465	2.362	0
3	2.396	4.400	0
4	1.388	1.850	0
5	3.064	3.005	0
6	7.627	2.759	1
7	5.533	2.088	1
8	6.922	1.771	1
9	8.675	− 0.242	1
10	7.673	3.508	1

• gradient des
→ assumption – error (log accuracy)

145 413

→ logistic function

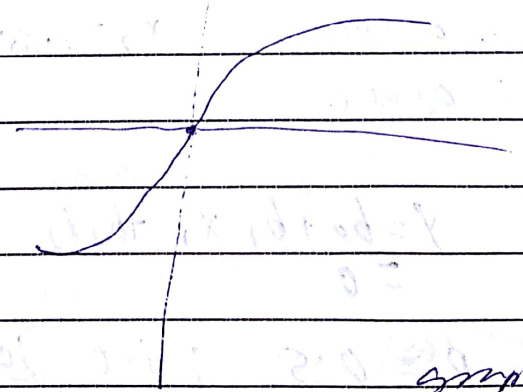
transformed = 1 / (1 + e ^ -x)

Euler Numerical Constant = 2.718

• accuracy of the model is to be checked in the end.

logistic function.

x	transformed value.
−5	0.00669
−4	0.01798
−3	0.04742
−2	0.11920
−1	0.2689
0	0.5
1	0.73105
2	0.8807
3	0.95257
4	0.98201
5	0.9933



graph of logistic function

Date
Page No.

Logistic Regression Model

$$\underbrace{y}_{\text{probability}} = \underbrace{b_0} + \underbrace{b_1} \cdot x_1 + \underbrace{b_2} \cdot x_2$$

probability

~~[illegible]~~

values to be found to find the best probability.

* Calculate a prediction using the current values of the coefficients.

* Calculate new coefficient values based on the error in the prediction.

Calculate Prediction

$$p = \frac{1}{(1 + e^{\wedge}(-\text{output}))}$$

Continue this and make graph of error.

for example,

let initial values of coefficients be

$B_0 = 0.000$ $x_1 = 2.781$
$B_1 = 0.000$ $x_2 = 2.5501$ $\rightarrow y \; \boxed{0}$
$B_2 = 0.000$

$$y = b_0 + b_1 x_1 + b_2 x_2$$
$$= 0$$

$p = 0.5$; i.e less than $= 0.5$ it will fall in 0 category alpha is a learning rate to be decided in the beginning, here let 0.3

Charlie

Date

Page No.

New co-efficients

$b = b + alpha * (y - p) * p * (1 - p) * x$ 1.0

(according to gradient descent) y-intercept.

$b_0 = 0.0 + 0.3(0 - 0.5) * 0.5 * (1 - 0.5) * 1$

$= 0 + 0.3(-0.5) * 0.5 * (0.5) * 1$

$= -0.0375$

$b_1 = 0.0 + 0.3 * (0 - 0.5) * 0.5 * (1 - 0.5) * 2.781$

$= -0.1042875$

$b_2 = 0.0 + 0.3 * (0 - 0.5) * (0.5) * (1 - 0.5) * 2.550$

$= -0.095625$

$y = -0.0375 + -0.1042875(2.781)$
$+ -0.095625(2.550)$

$= -0.0375 - 0.29 - 0.2438$

$= -0.57$

$\boxed{p = 0.361}$ i.e less than 0.5

so it will fall in 0 category

· logistic regression.

Date
Page No.

Performance of Logistic Regression

1. AIC (Alkaline Information Criteria)

2. NULL Deviance & Residual Deviance

3. Confusion Matrix] True Positive
True Negative

HYPOTHESIS TESTING

When one talks about hypothesis, one simply means an assumption or some supposition to be proved or disproved.

→ A new medicine, you think might work.

→ A possible location of new species.

NULL hypothersis:
Actual

NULL hypothesis & Alternative (Actual) hypothesis.

[p-value].

Charlie
Date
Page No.

eg: If we are to compare method A with method B about its superiority, and if we proceed on the assumption that both methods are equally good.

Then this assumption is termed as Null Hypo.
$\boxed{H_0}$

As against this, we may think that the method A is superior, or the method B is inferior.
This is termed as Alternative Hypothesis.
$\boxed{H_a}$

$H_0 = \mu = 100 \quad = 100$

H_a ————————— 105

· population data.

The Level of Significance

The level of significance is an important constant in the hypothesis testing.
It is always some % which should be chosen with great care, thought and reason.

In case we take significance level at 5%, then this implies H_0 [Null hypothesis] will be rejected, when the sampling result has a less than .05 probability of ~~occurring~~ occuring if H_0 is true.

Scanned by CamScanner

Decision or Test of Hypothesis

	Decision	
	Accept H_0	Reject H_0
H_0 (true)	Correct decission	Type I error (α error)
H_0 (false)	Type II error (β error)	Correct Decission.

Two tailed & One tailed test