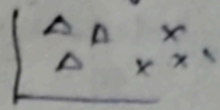
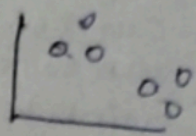
→ classification
→ Prediction (Regression). } supervised learning (x, y)



→ clustering.
→ Dimensionality Reduction. } unsupervised learning. (x)



Classification

— Email classification

Email text (x) → Rules → spam / Not spam (y)

Data. ↑ (into Rules)

(y) ⇓ categorical variable (discrete)

~~Data (x)~~ ~~label (y)~~

~~Email~~

Data (x)	y
Email	label
Meeting at 10.00pm	Not spam
won $10000	spam
you won lottery	?

Train: Email, Meeting at 10.00pm, won $10000 rows
Test: you won lottery

Steps.
1. Train
2. Test
3. Deploy the model.

Regression

1. Weather prediction.

Previous day temp, humidity, → [ ] → Temp

Y (continuous).

→ classification Y is categorical / discrete.

→ Regression Y is continuous.

→ clustering No Y.

→ Dimensionality Reduction X.

(2) Score prediction in cricket.

Current run rate, wickets → $\boxed{}$ → final score.

~~(Y continuous)~~

(3) Stock Market price prediction.

(1) Supervised Learning (x,y)

Ex: (1)

(1) Learning from solved Examples – Training

(2) Exercises – validation.

(3) Exam – ~~Textice~~ Testing.

X	Y → Human knowledge (supervision)
Email	label
Meeting at 10	
won lottery	

Ex: (1)

Face Recognition is a supervised.

Ex: (2)

card Fraud Detection

Ex: (3)

Disease prediction

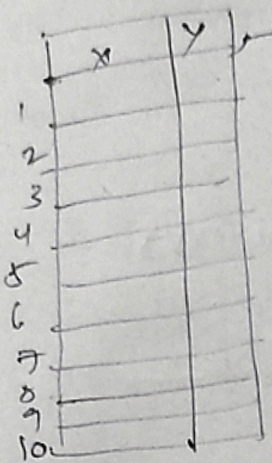
Symptom(x) → $\boxed{}$ — Disease (Y)

(4) Stock Price prediction.

Previous 10 days price $\rightarrow \square \rightarrow$ Tomorrow price (Y)
(X)

Dataset

→ split the data into trainset, validation set, test set.



→ Train (60%) 80%
validation (20%) 10%
Test (20%) 10%

Select by Randomization.

Train - 5, 10, 1, 6, 2, 7, 4, 3
Validation - 9
Test - 8

Model

(X, Y) → true label

$X \rightarrow \boxed{F(x)} \rightarrow \hat{y} \approx y$ ← predicted label

model → $\boxed{\hat{y} = f(x)}$

y = true age
$\hat{y}$ = predicted age.

Regression

	No. of hours	Marks
train	1	50
	2	55
	3	65
	4	70
	5	80
Test	3.5	?

Linear function

(Line)

① $\boxed{y = mx + c}$ → simplest function to convert x into y.
(O/P: y; I/P: x; m: slope; c: intercept)

② $y = x_1 + x_2$

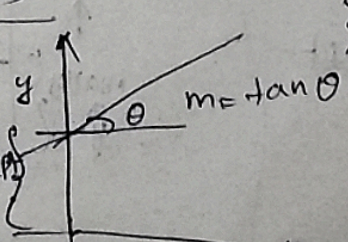
Non linear function

① $y = x^2$
② $y = x_1^2 + x_2^2$

linear function

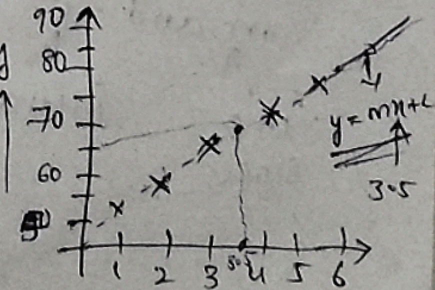
$y = mx + c$

m and c we can control to define the line



$m = \tan\theta$

Scatter plot



$$\boxed{\hat{y} = f(x) = mx + c}$$

Parameters

- controls the model

$y = mx + c$

m and c are the parameters (2 parameters)

Training = finding right parameters.

Deep Learning
↓
Neural Network
↓
Implied by Brain
↓
Neurons.

$$y = \beta_0 + \beta_1 x.$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Model

$$\hat{y} = f(x)$$

example

$$\hat{y} = mx + c$$

parameters

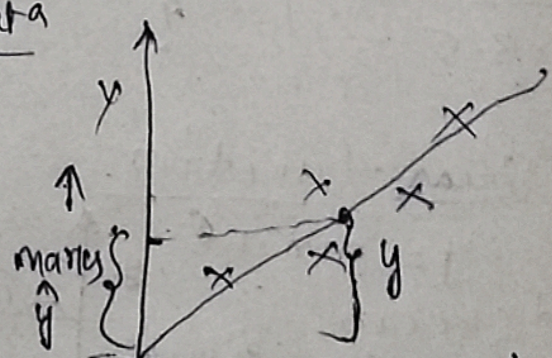
m and c
↑ tan θ slope
↑ intercept.

Train- finding the right parameter values.

Simple linear Regression :-

$$\hat{y} = \theta_0 + \theta_1 x$$

Data



$y = 4 + 10x$

$x = 5, \hat{y} = 54$

$y = 60$

Error $= y - \hat{y}$

Model   model 1   model 2

X	Y	$\hat{y}$	ε_i	$\hat{y}$	ε_i
1	50	14	36	42	8
2	55	24	31	44	11
3	65	34	31	46	19
4	70	44	26	48	22
5	80	54	26	50	30

Total error.

$\hat{y} = 4 + 10x$

$x = 5, \hat{y} = 54$

$SSE = \sum_{i=1}^{n} (y_i - \hat{y})$

$= \underline{90}$ (Error).

$\hat{y} = 4 + 10x$ $SSE = \sum_{i=1}^{n} (y_i - \hat{y})$

$\hat{y} = 40 + 2x$ $= \underline{150}$.

$\hat{y} = 41.5 + 7.5x$

Error

model 3

$\hat{y}$	ε_i
49	1
56.5	1.5
64	1
71.5	1.5
79	1
0	6

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

⇒ Sum of squared error.

(1) All errors will become positive.

(2) It magnifies large error.

Finding the best parameters (minimum Error).

(1) using formula:

$$\theta_0 = \bar{y} - \theta_1 \bar{x} = 41.5$$

$$\theta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$= 7.5$

└ parameters

(2) Error

(3) Train

└ Find the parameter for which error is minimum.

<u>Regression</u>

x	y
1	50
2	55
3	65
4	70
5	80

$\hat{y} = \theta_0 + \theta_1 x$

Find θ_0 and θ_1 such that the error is minimum.

$y_i - \hat{y}_i$

$|y_i - \hat{y}_i|$

$SSE = \sum_{i=1} (y_i - \hat{y}_i)^2$

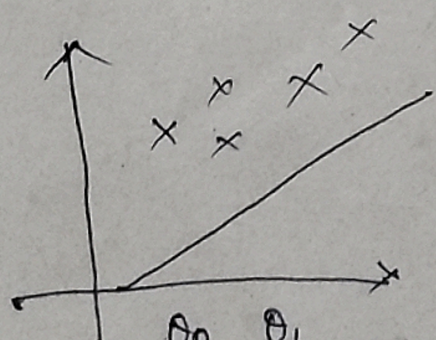
$= \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2$

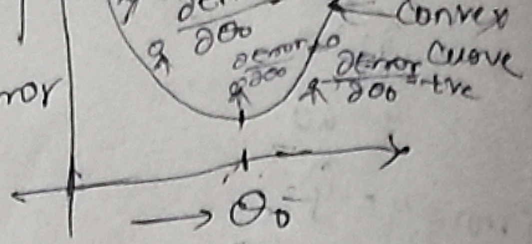
$= \sum_{i=1}^{n} x^2 \to 2x$

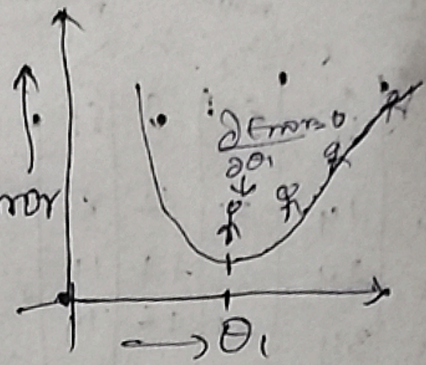
<u>Formula</u> :

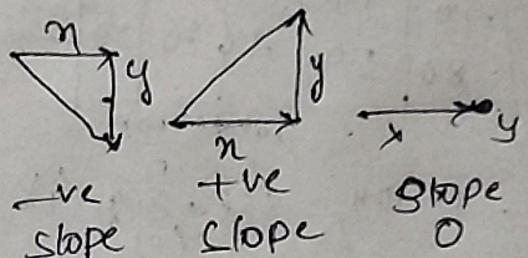
$\theta_0 = \bar{y} - \theta_1 \bar{x}$

$$\theta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$









$\frac{dy}{dx} \to$ Derivation

change in x represents change in y.

$\frac{\partial Error}{\partial \theta_0} = 0 \quad \frac{\partial Error}{\partial \theta_1} = 0$

<u>Find θ_0</u>

$\frac{\partial SSE}{\partial \theta_0} = 0$.

$\Rightarrow 2 \sum (y_i - \theta_0 - \theta_1 x_i)(-1) = 0$

$\Rightarrow -2 \sum (y_i - \theta_0 - \theta_1 x_i) = 0$

$\Rightarrow \sum (y_i - \theta_0 - \theta_1 x_i) = 0$

$\Rightarrow \sum y_i - \sum \theta_0 - \sum \theta_1 x_i = 0$

$\Rightarrow \sum y_i - n\theta_0 - \theta_1 \sum x_i = 0$

$\Rightarrow n\theta_0 = \sum y_i - \theta_1 \sum x_i$

$\Rightarrow \theta_0 = \frac{\sum y_i - \theta_1 \sum x_i}{n}$

$\Rightarrow \theta_0 = \frac{\sum y_i}{n} - \frac{\theta_1 \sum x_i}{n}$

$\Rightarrow \boxed{\theta_0 = \bar{y} - \theta_1 \bar{x}}$

$$\frac{\partial SSE}{\partial \theta_1} = 0$$

$$\Rightarrow 2\sum(y_i - \theta_0 - \theta_1 x_i)(-x_i) = 0$$

$$\Rightarrow -2\sum(y_i - \theta_0 - \theta_1 x_i)(x_i) = 0$$

$$\Rightarrow \sum(y_i - \theta_0 - \theta_1 x_i)x_i = 0$$

$$\Rightarrow \cancel{\sum(x_i y_i - \theta_0 x_i)x_i = 0}$$

$$\Rightarrow \sum(x_i y_i - \theta_0 x_i - \theta_1 x_i^2) = 0$$

$$\Rightarrow \sum x_i y_i - \theta_0 \sum x_i - \theta_1 \sum x_i^2 = 0$$

$$\Rightarrow \sum x_i y_i - \left(\frac{\sum y_i - \theta_1 \sum x_i}{n}\right)\sum x_i - \theta_1 \sum x_i^2 = 0$$

$$\Rightarrow n\sum x_i y_i - (\sum y_i - \theta_1 \sum x_i)\sum x_i - n\theta \sum x_i^2 = 0$$

$$\Rightarrow n\sum x_i y_i - \sum x_i \sum y_i + \theta_1 (\sum x_i)^2 - n\theta_1 \sum x_i^2 = 0$$

$$\Rightarrow n\sum x_i y_i - \sum x_i \sum y_i = \theta_1 (n\sum x_i^2 - (\sum x_i)^2)$$

$$\Rightarrow \boxed{\theta_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}}$$

Regression :-

- Estimating θ_0 and θ_1 using formula.

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

- Estimating $\theta = \{\theta_0, \theta_1\}$ using normal form

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

1.	50
2	55
3	65
4	70
5	80.

$$y = \begin{bmatrix} 50 \\ 55 \\ 65 \\ 70 \\ 80 \end{bmatrix}$$

$$\hat{y}_i = \theta_0 + \theta_1 x_i = \theta_0 \cdot 1 + \theta_1 \cdot x.$$

$$= \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}}_{x} \times \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}_{\theta}$$

$$\theta = \underbrace{(x^T x)^{-1}}_{2\times 2} \cdot \underbrace{x^T y}_{2\times 1} \Rightarrow 2\times5 . 5\times2 = 2\times2 \quad 2\times5 . 5\times1 = 2\times1$$

$$\hat{y} = x\theta$$

$$x = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}_{5\times 2} \quad y = \begin{bmatrix} 50 \\ 55 \\ 65 \\ 70 \\ 80 \end{bmatrix}_{5\times 1}$$

Python :- import numpy as np.

```
x = np.array((1,1),(1,2),(1,3),(1,4),(1,5))
y = np.array(50, 55, 65, 70, 80)
theta = np.linalg.inv(x.T @ x) @ x.T @ y
print(theta).
```

$$\theta_0 = \bar{y} - \theta_1 \bar{x} = 41.5$$

$$\theta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 7.5$$

1	50	49
2	55	56.5
3	65	64
4	70	71.5
5	80	79

MLR : $\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

Polynomial Regression :

$\hat{y} = \theta_0 + \theta_1 x + \theta_1 x^2$

↑ Polynomial feature

x	x^2	y
1	1	
2	4	
3	9	
4	16	
5	25	

$SSE = \sum_{i=1}^{n} (y_p - \hat{y}_p)$

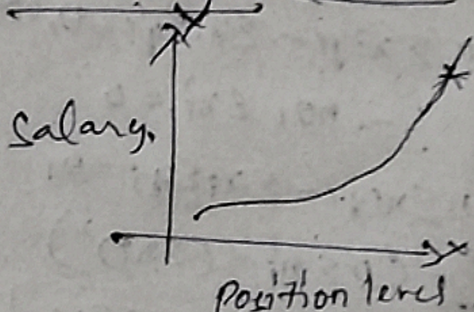
$= \underline{(y - \hat{y})^T (y - \hat{y})}$

Ex

$y = \begin{bmatrix} 50 \\ 55 \\ 65 \\ 76 \\ 80 \end{bmatrix} \quad \hat{y} = \begin{bmatrix} 49 \\ 56.5 \\ 64 \\ 71.5 \\ 79 \end{bmatrix}$

$y - \hat{y} = \begin{bmatrix} 1 \\ -1.5 \\ 1 \\ -1.5 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ -1.5 \\ 1 \\ -1.5 \\ 1 \end{bmatrix}$

Non linear relation



$= y^T y - y^T X\theta - \theta^T X^T y + \theta^T X^T X\theta$

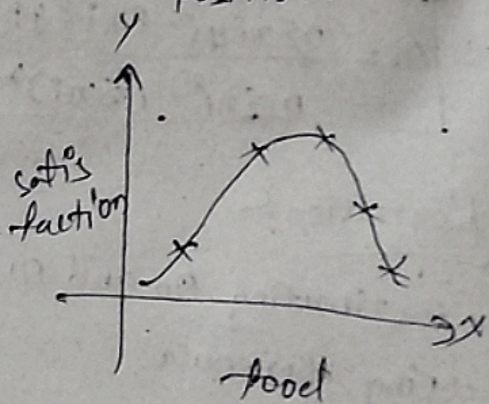
$= 1\times5 \cdot 5\times2 \cdot 2\times1 \mid 1\times2 \cdot 2\times5 \cdot 5\times1$

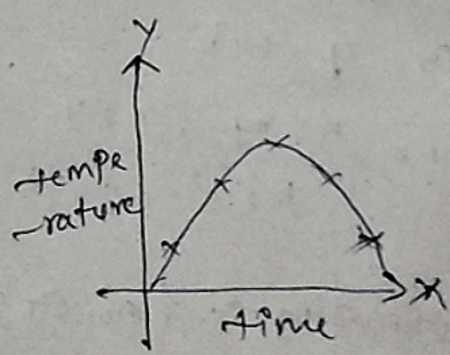
$= 1\times1 \mid = 1\times1$

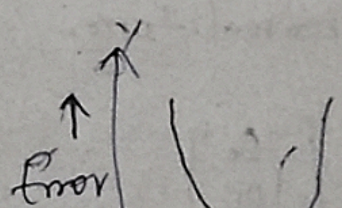
$\boxed{y^T X\theta = \theta^T X^T y}$

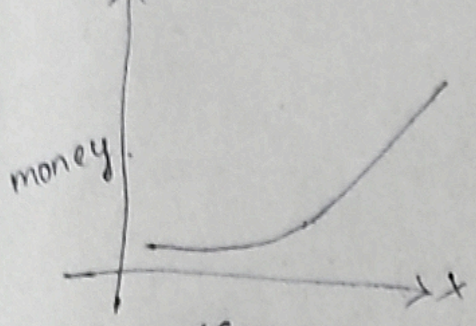
$= y^T y - \theta^T X^T y - \theta^T X^T y + \theta^T X^T X\theta$

$= y^T y - 2\theta^T X^T y + \theta^T X^T X\theta$







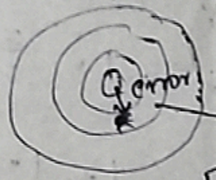


Evaluation metric

→ What is the performance on test data ?

X	Y	
		} train ↳ Error → $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)$
		$MSE = \frac{SSE}{n}$
		$RMSE = \sqrt{MSE}$
		} Test → Evaluation metric ↳ R^2 → Adjusted R^2

Archery



→ score ↓ Evaluation metric.

$\hat{y} = f(x)$.

$$\text{Variance} = \frac{\sum (y_i - \bar{y})^2}{n}$$

$1 - 3 = -2$
$2 - 3 = -1$
$4 - 3 = 1$
$5 - 3 = 2$

$\frac{12}{4} = 3$

↓ distances from mean.

Explained variance.

$$R^2 = \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

↓ Variance in data

features ↓

X	y	$\hat{y}$	$y - \hat{y}$	$(y-\hat{y})^2$
	50	49.5	1	1
	55	56.5	-1.5	2.25
	65	64	1	1
	70	71.5	-1.5	2.25
	80	79	1	1
	64	64.1		7.5

$$= \frac{(50-49.5)^2 + (55-56.5)^2 + (65-64)^2 + (70-71.5)^2 + (80-79)^2}{(50-64)^2 + (55-64)^2 + (65-64)^2}$$

$$= \frac{6.25 + 2.25 + 1 + 2.25 + 1}{196 + 81 + 1 + 36 + 256}$$

$= 6.75$

$Y-\bar{Y}$	$(Y-\hat{Y})^2$
-14	196
-9	81
1	1
6	36
16	256
	570

$$R^2 = \frac{7.5}{570} = 0.013$$ → Near to 1 Good fit

Near to 0 Bad fit.

$$\text{Adjusted } R^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-k-1}\right)$$

k = no. of features
n = no. of data points

If you add unnecessary feature Adjusted R^2 will be less.

k = 1
n = 5

$$= 1 - \left(\frac{(1-0.013)(4)}{5-1-1}\right) = 1 - \frac{3.948}{3}$$

$$= 1 - 1.316$$

$$= -0.316$$

Classification

- Email classification.

email x → $f(x)$ → Spam-1 / Not spam-0

class label y $\hat{y}$ Discrete

x	y	$\hat{y}$	
	1	1	✓
	0	1	×
	1	0	×
	1	1	✓
	0	1	×

Classification Evaluation metrics

Classification matrix

Actual \ Predicted	1	0
1	TP 3	FN 1
0	FP 3	TN 1

TP = True Positive.
TN = True Negative.
FP = False Positive
FN = False Negative.

y	$\hat{y}$
1	1
0	1
1	0
1	1
0	1
1	1
0	1
0	0

Decision Tree classifier :-

To measure uncertainity $\{P_1, P_2\}$

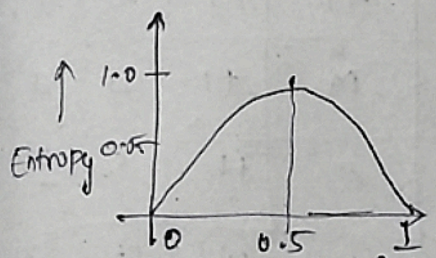
$\{50\%, 50\%\} \rightarrow$ more uncertainity

$\{10, 10\} \rightarrow$ more impurity

Entropy $= -P_1 \log_2 P_1 - P_2 \log_2 P_2$

$\{90\%, 10\%\} \rightarrow$ less uncertainity

$\{90, 10\} \rightarrow$ less impurity



$\{0,1\} \longrightarrow P_1$ $\{1, 0\}$

$\{P_1, P_2\}$

Information Gain

Genere	time	watch movie ?
~~A~~	E	Y
C	M	N
A	M	Y
D	E	N
C	E	N

x_1	x_2	x_3	x_4	label

$E(x_1)$
$E(x_2)$
$E(x_3)$
$E(x_4)$

E(root) = 0.97.

{Y:2, N:0} {Y:0, N:2} {Y:0, N:1} splits.

2/5 2/5 1/5

= 0 = 0 = 0

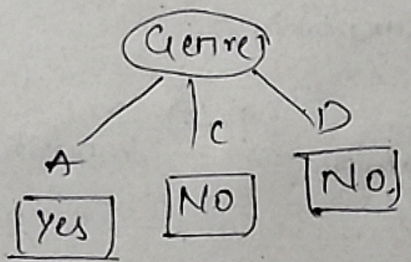
Average Entropy = $\frac{2}{5} \times 0 + \frac{2}{5} \times 0 + \frac{1}{5} \times 0$

= 0.

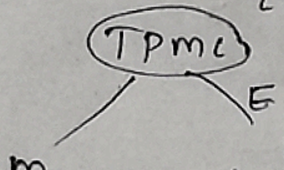
IG (Entropy) = Reduction is uncentainity/

impurity = 0.97 − 0 = 0.97 ✓
best to take

Decision Tree :



{Y:2, N:3}.



{Y:1, N:1} {Y:1, N:2}

2/5 3/5

En (Timem) E (Time E)

= 1 = 0.91

Average Entropy = $\frac{2}{5} \times 1 + \frac{3}{5} \times 0.91$
(of children) = 0.946

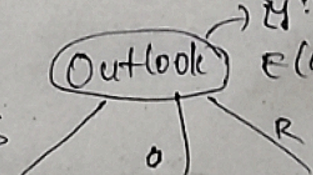
IG (Time) = 0.97 − 0.946
= 0.024

Day	Outlook	Temp	Humidity	Wind	Play Tennis
1	S	H	H	W	N
2	S	H	H	S	N
3	O	H	H	W	Y
4	R	M	H	W	Y
5	R	C	N	W	Y
6	R	C	N	S	N
7	O	C	N	S	Y
8	S	M	H	W	N
9	S	C	N	W	Y
10	R	M	N	W	Y
11	S	M	N	S	Y
12	O	H	H	S	Y
13	O	H	N	W	Y
14	R	M	H	S	N
15	S	C	H	S	?

Sunny Hot High, Strong Yes
overcast mild Normal weak No
Rainy cool

①

{Y:9, N:5}



E(outlook) = 0.94

{Y:2, N:3} {Y:4, N:0} {Y:3, N:2}
5/14 4/14 5/14

E(outlook$_S$) E(outlook$_O$) E(outlook$_R$)
= 0.97 = 0 = 0.97

Average Entropy = $\frac{5}{14} \times 0.97 + \frac{4}{14} \times 0 +$
$5/14 \times 0.97 = 0.69$.

IG (outlook) = 0.94 − 0.69 = 0.24

H   m

{Y:3,N:2}   {Y:3,N:2}   {Y:3,N:1}

$E(Temp_H)$ $E(Temp_m)$ $E(Temp_c)$

$= 0.97$ $= 0.97$ $= 0.81$

$$\text{Average Entropy} = \frac{5}{14} \times 0.97 + \frac{5}{14} \times 0.97 + \frac{4}{4} \times 0.81$$

$$= 0.924.$$

$$IG(Temp) = 0.94 - 0.924$$

$$= 0.016.$$

(3) Humidity

{Y:3, N:4}   {Y:6, N:1}

0.98   0.98

$$\text{Average Entropy} = \frac{7}{14} \times 0.98 + \frac{7}{14} \times 0.98.$$

=

$IG(Temp) = 0.94 -$

=

Wind

S   W

{Y: , N: }   {Y: , N: }