CLASSIFICATION, REGRESSION

&

ACCURACY METRICS

8 Jan 2019

Features

O/P (Predict)

Color	Size	Texture	Condition
			Good
			Bad
			good

Manually

SUPERVISED
LEARNING

Training set

X
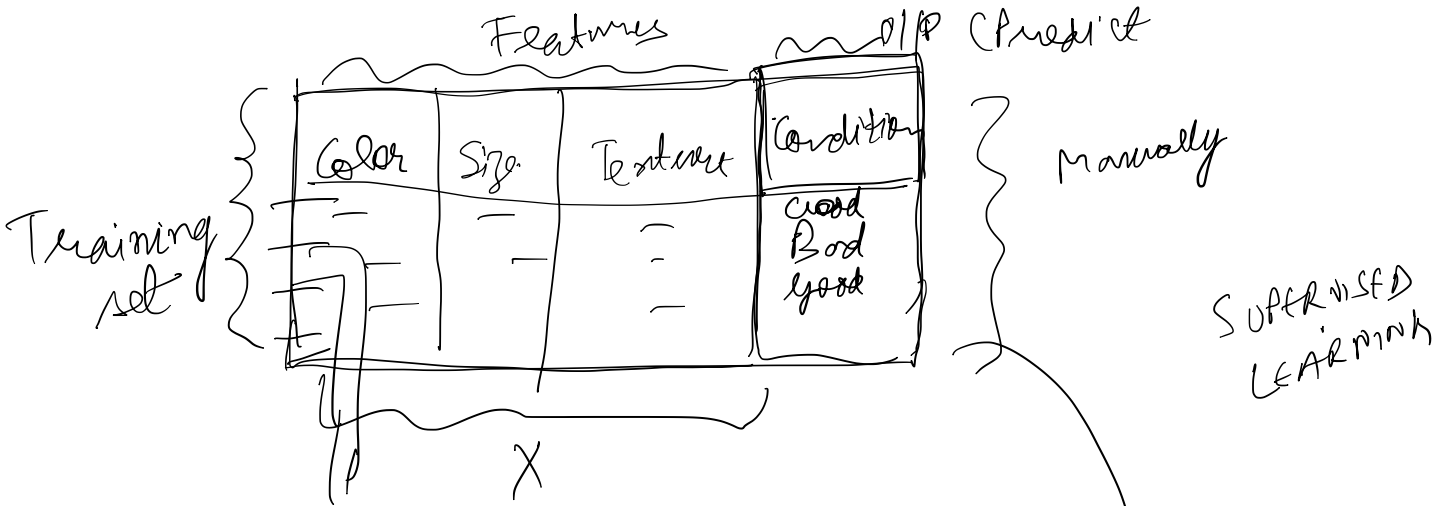
Rows of X indicate sample
Columns " " features.

LEARN

'Condition' = f(Color,
Size, Texture, ...)

New tomato

Test set

Size	Color	Texture	Condition
✓	✓	✓	?
✓	✓	✓	?
✓	✓	✓	?

Condition is discrete (Classification)

Prediction ($\hat{y}$)

$$\begin{bmatrix} Good \\ Bad \\ Good \\ Good \\ Bad \end{bmatrix}$$

Ground Truth

$$\begin{bmatrix} G \\ B \\ B \\ B \\ B \end{bmatrix}$$

Prediction ($\hat{y}$)

Good
Bad
Good
Good
Bad

Ground Truth

G
B
B
B
B

$$Accuracy = \frac{Count(\hat{y} = y)}{Size\ of\ \hat{y}\ or\ y}$$

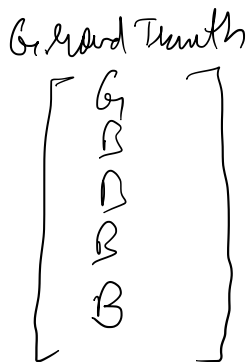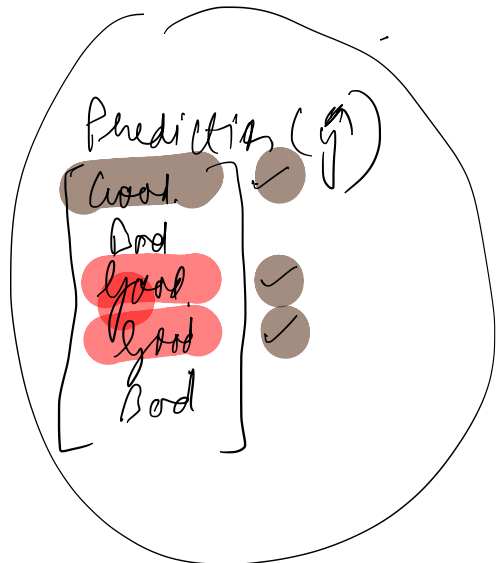$$= \frac{3}{5} \times 100\% = 60\%$$

INBALANCED CLASSES

NO
⋮
NO

Ground Truth

NO Cancer
Yes
NO
NO
⋮
NO

} 100 Samples

Accuracy = 99%

Prediction (ŷ)

Good.
Bad
Good.
Good
Bad

Ground Truth

$$\begin{bmatrix} G \\ B \\ B \\ B \\ B \end{bmatrix}$$

$$\begin{bmatrix} G \\ B \\ B \\ B \end{bmatrix} \xrightarrow{\hat{y}} \begin{bmatrix} G, T. \\ G \\ G \\ B \\ G \end{bmatrix}$$

Precision $= \dfrac{1}{3}$

$$= \dfrac{count(\hat{y} = y = Good)}{count(\hat{y} = Good)}$$

Recall
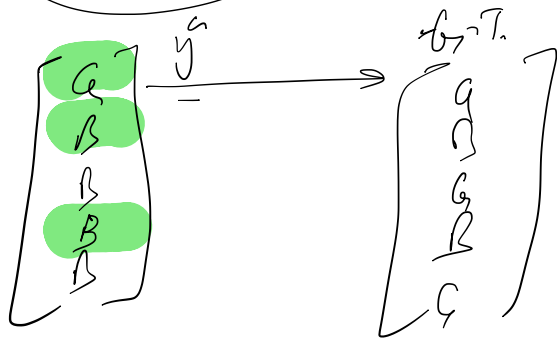
Prediction ($\hat{y}$)

- Good
- Bad
- Good
- Good
- Bad

Ground Truth

- G
- B
- G
- B
- G

$$Recall = \frac{2}{3} = \sim 67\%$$

$$= \frac{count(y = \hat{y} = Good)}{count(y = Good)}$$

Cancer Examples $\hat{y}$

- NO
- Yes
- |
- |
- |
- NO

G.T. (y)

- NO
- NO
- ...
- Yes
- ...
- NO

$$Accuracy = \frac{98}{100} \times 100\% = 98\%$$

$$Precision = 0/1 = 0\%$$

$$Recall = 0/1 = 0\%$$

Confusion Matrix

	G.T. cancer	G.T. (No cancer)
Pred Cancer	0	1
Pred No cancer	1	98

TRUE POSITIVE

FALSE POSITIVE

TRUE NEGATIVE

FALSE NEGATIVE
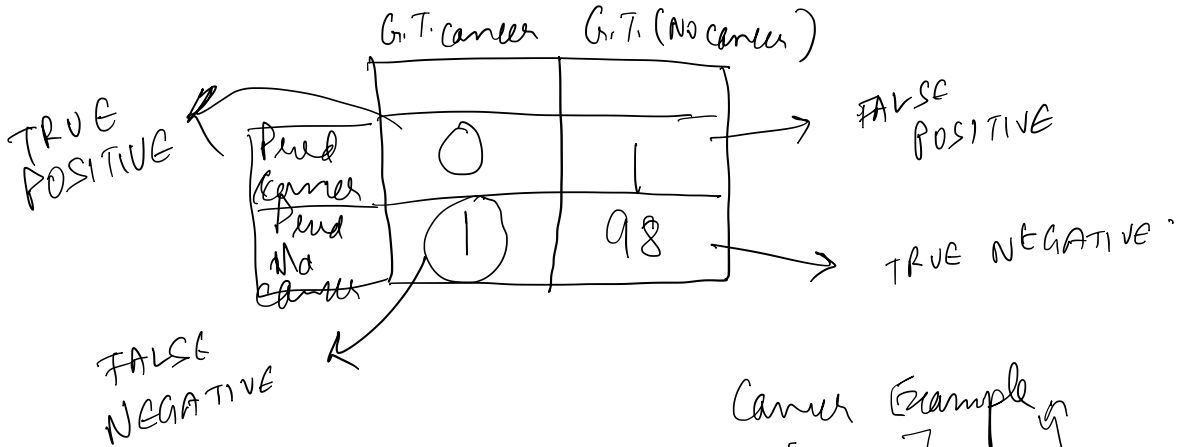
$$PRECISION = \frac{TP}{TP + FP}$$

$$RECALL = \frac{TP}{TP + FN}$$

Cancer Example

G

No
Yes
⋮
No

NO
NO
⋮
Yes
⋮
NO
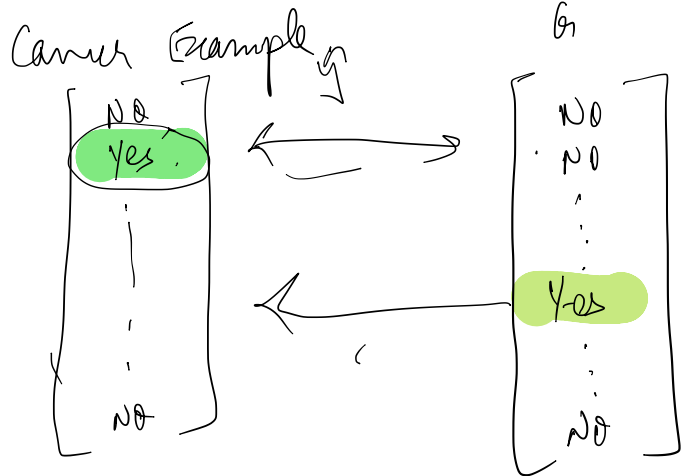
F-Score $= \dfrac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ (0 and 1)

$$\begin{array}{|c|c|} \hline 90 & 4 \\ \hline 1 & 1 \\ \hline \end{array}$$

TP = 90, FP = 4, TN = 1, FN = 1

Precision
$\left(\dfrac{90}{94}\right)$,   Recall $\left(\dfrac{90}{91}\right)$ Q. Precision, Recall, F-Score

F-Score = $\boxed{.9523}$ → Random
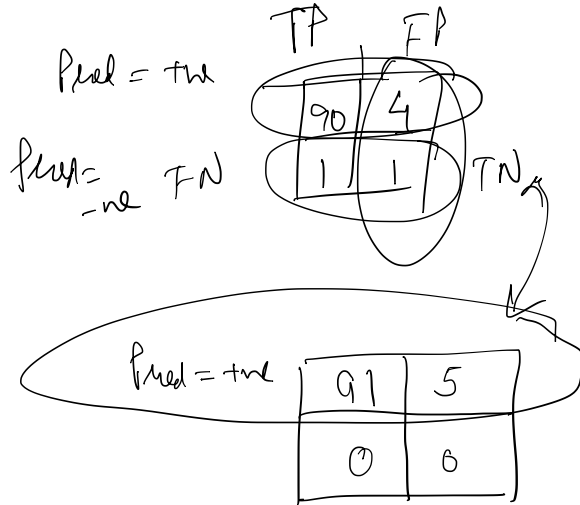
Matthews Correlation Coefficient

$(-1 \text{ to } 1)$ → Highly correlated (Accurate)

$$\dfrac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

= $\boxed{0.14}$

G.T=+ve   G.T=Neg.

TP   FP
Pred = +ve

90	4	

Pred = -ve   FN

		TN
1	1	

Pred = +ve

91	5
0	0

96 samples
of have negative
class

$$\frac{91}{96}$$

Regression

Lat	Long	Altitude		Season
				Rainfall (mm/year)

Lecture Room	# Students	T. of day		Energy (kWH)
				(Projector + AC + Lights)

Predicted ($\hat{y}$)	G.T (y)
10	20
20	20
30	40
40	50

$$\text{Mean Squared Error} = \frac{\sum(\hat{y}_i - y_i)^2}{N}$$

(MSE)

$$\text{Root Mean Squared Error} = \sqrt{MSE}$$

Predicted ($\hat{y}$)	G.T (y)
10	0
0	10

$$\text{Mean Absolute Error} = \frac{\sum(y_i - \hat{y}_i)}{N}$$

(MAE)

$$MAE = \frac{|10-0| + |10-0|}{2} = 10$$

$$\text{Mean error} = \frac{10-0 + 0-10}{2} = 0$$