1

Lecture Notes

Prepared by: Kazi Junaid Mahmud

on

Machine Learning - Basic ML

LO1 - What is ML and Why ML

* What is ML and when to use ~~ML~~ ML?
  ⇒ a branch of AI and computer science,
  ⇒ Focusses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Child $\xrightarrow{7y}$ Young $\xrightarrow{20y}$ Adult $\xrightarrow{30y}$ Boomer

* information gained.
* experience gained.
  ↳ learned through the experiences.

model 1 (Algorithm) ⟶ model 2 ⟶ m3 ⟶ m4

* improvement through training.
* works like human, learns and improves.

(2)

1* Q

input x
if (x % 2) == 0
    "even"
else
    'odd'

* rules are defined,
* no confusion in results, if rules followed.

Now let's talk about a complicated problem

floor	Location	size	rent cost
7	Banani	1500	50000
4	Tongi	2200	55000
2	Banani	1200	30000
5	Dhanmondi	1500	40000

→ let's set a rule that if (size > 1500) then rent_cost = 40000.

* but for tongi, even though size is much bigger, the price isn't much higher a much smaller flat of $\boxed{\text{size } 1500}$ in Banani.

→ Also the same can be thought about the following cases.
- Location
- floor.

Hence, we can summarize that,

"We can not ~~deci~~ decide the rent cost based on just one feature."

③

let's create some complex rules

if (location point Banani flat 1200 floor 7)
40000

...
...
else if
else if
else

* this may seems to work primarily. But if we introduce new data, for example, another row added to the prev table:

floor	location	size	rent cost
3	Uttara	1800	??

this will not give accurate answer, as no condition for "Uttara" "3" and "1800" was found. hence this system will fail.

** So when we can not set the rules, we ~~us~~ use Machine Learning.

(4)

Lecture Notes

on

Machine Learning - Basic ML

L02 - Types of Machine Learning

Prepared by: Kazi Junaid Mahmud

ML

- Supervised
- Unsupervised
- Semi Supervised
- Reinforcement Learning.

→ When supervised, and when unsupervised?

- Supervised

cat dog classification

image 1	cat
image 2	dog
image 3	cat
image 4	dog

↓ label

* we will take an image of either a cat or dog and the model will predict whether image is of a cat or a dog.

* How do the algorithm understands that this picture is a cats picture? or a dogs picture?

→ Based on $\boxed{\text{features}}$ / some patterns

5

Label:

A picture has the name cat / dog. So, it is already given for the training dataset.

The classes (or labels) are specifically mentioned in supervised learning.

Unsupervised Learning:

The labels will not be given. Just the data.

	height
img 1	f1
img 2	f2
img 3	f3
img 4	f4

the data will be ~~classified~~ grouped based on similarities in features. But the labels are not mentioned.

Supervised Learning
- → Regression → continuous values
- → Classification → discrete values

Regression:

- Any value within a range.
- Example. CGPA, Rent cost
- Continuous Values
- Target not known exactly.

Classification

- Any value of a list of discrete values.
- Pass/fail, cat/dog.
- Discrete Values.
- Target known (the options)

6

Lecture Notes
on
Machine Learning - Basic ML

L03 - Linear Regression Part - 1

prepared by: Kazi Junaid

Regression

size	rent cost
1200	25000
1600	30000
1800	35000
2000	??

→ what will be the rent cost when the size is 2000 ?

Linear Regression:

$f(x) = 2x + 3$

$f(x)$ → dependent (target)

x → independent (feature)

$$\boxed{\text{rent\_cost} = \underbrace{20}_{\text{guess}} \times \text{size}}$$ ⇒ hypothesis

⇓

$$\boxed{R = 20S}$$

- we randomly guessed the parameter.

Prediction: size from the dataset

Prediction	Actual
$R = 20 \times 1200 = 24000$	25000
$R = 20 \times 1500 = 30000$	30000
$R = 20 \times 1800 = 36000$	35000

* how do I know the performance of my model?

$$Loss = \sum \left| Actual - prediction \right|$$

$$= |-1000| + 0 + |1000|$$

$$= 2000$$

total Loss $= 2000$

So, The less "Total Loss" is, higher the "Accuracy"

Loss $= 2000$ (3 samples)

- But if there are 1000 samples.

	sample	error for each	ΣL
m1	3	1000	3000
m2	1000	5	5000

the second model (m2) is better, as it has lower loss for one data point.

so, instead of "total loss" we use "avg loss".

$$Loss = \frac{1}{N} \sum \left| Actual - Prediction \right|$$

↑ number of samples

m1 → Loss = 3000 → Avg Loss = ~~1000~~ 1000 ✗

m2 → Loss = 5000 → Avg Loss = 5 ✓

$\boxed{\text{m2 = better}}$

State of the Art:

2018 → Loss = 1000 ✗

2020 → Loss = 100 ✗

2024 → Loss = $\boxed{10}$

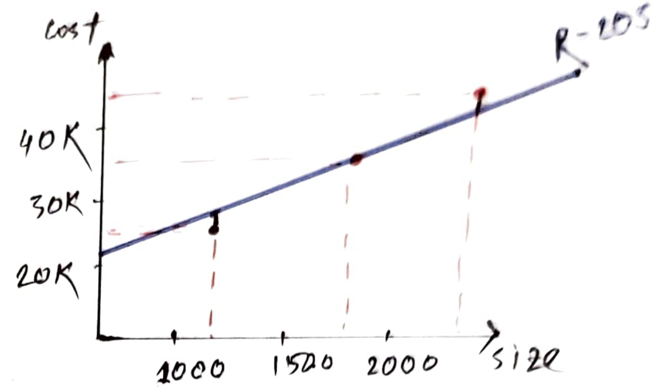
this is the current "state of the art"

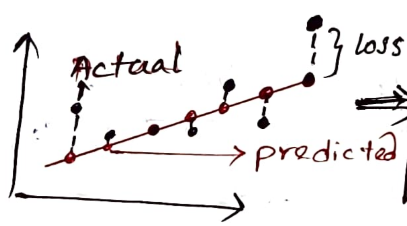
R = 20 S → 20k 2000 → 40000

$\boxed{R = 20S}$ → is the model, as we are happy with the result.

L04 - Linear Regression Part 2

Linear ?	
$y = mx + c$	✓
$y^2 = mx + c$	✗
$y^{1/2} = mx + c$	✗





⇒ not all the points are on the line.

- All points can not be "fit" on the line.

$\min(\text{Loss})$ $R = 205$ | $y = wx$ — y ⇑ target, x ⟶ feature, w ⇓ weight

Gradient Descent:

$$W_{new} = W_{old} - \underbrace{\alpha}_{\text{learning rate}} \frac{\partial}{\partial w}\left(\text{Loss}\right)$$

L05 - Linear Regression Evaluation Metrics

	size	price	prediction
1	1000	20000	21000
2	1200	25000	27000
3	1600	35000	31000

y $\hat{y}$

Mean Absolute Error:

$$MAE = \sum \frac{1}{N} |y - \hat{y}|$$

by applying here : . $MAE = \frac{1000 + 2000 + 4000}{3} = \frac{7000}{3}$

Mean Squared Error:

$$MSE = \Sigma \frac{1}{N} (y - \hat{y})^2$$

$(unit)^2$

Root Mean Squared Error:

$$RMSE = \sqrt{MSE}$$

$$= \sqrt{\Sigma \frac{1}{N} (y - \hat{y})^2}$$

unit

- This one is vastly used in cases of Regression

R^2 (co-efficient of Determination):

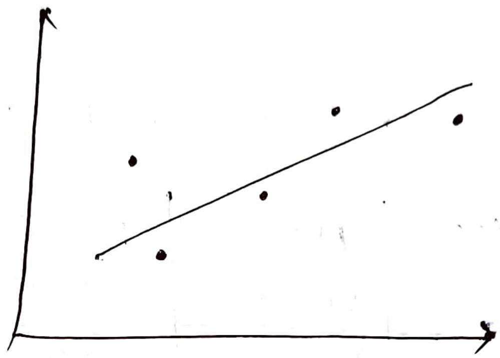


fig 1

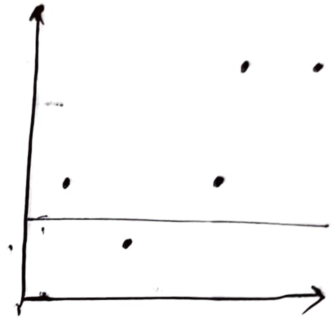


fig 2

$0 \leq R^2 \leq 1$

- R^2 যত 1 এর কাছাকাছি যাবে, বুঝা যাবে লাইন টা point এর তত কাছে দিয়ে যাচ্ছে. ও যত 0 এর কাছে, তত বাজে

$$R^2 = 1 - \frac{\Sigma (y - \hat{y})^2}{(y - \bar{y})^2}$$

y : Actual

$\hat{y}$: Prediction

$\bar{y}$: avg

LOG. Classification Problem

	Weather	Outlook	Play golf?
1	hot	sunny	yes
2	cold	sunny	yes
3	cold	Rainy	No
4	hot	sunny	Yes
5	Cold	sunny	Yes

☐ **Zero R classifier:**

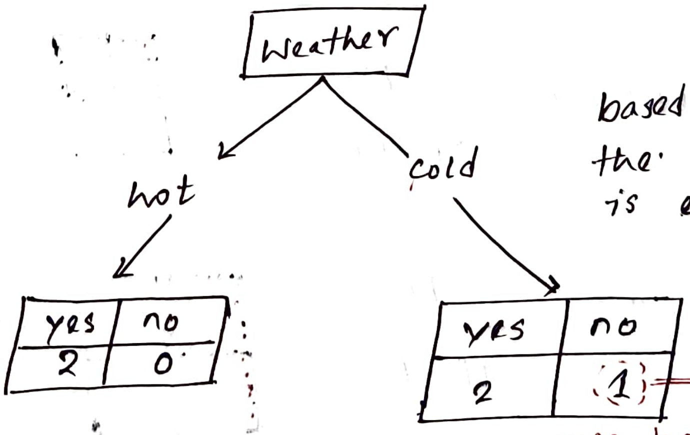
majority count → yes (4) ✓ → no - (1)

prediction → Yes

যা বেশিবার yes বলবে.

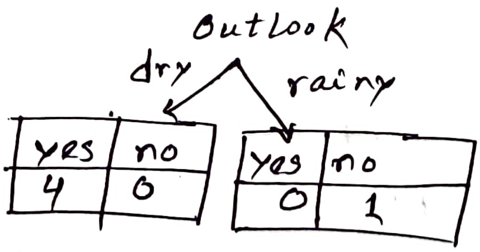
☐ **One R classifier:**

* will consider one feature, for prediction



based on weather the outcome is examined.

→ error! We have both yes and no with result > 0



the same is done in case of outlook

* We will try generate a general formula.

→ since $\boxed{\text{outLook}}$ has no conflict, it will be our model. ✓

L 07 - classification Evaluation part 1

Accuracy, Precision, Recall, F1

Confusion Matrix

positive - (good) P
negative - (bad) N

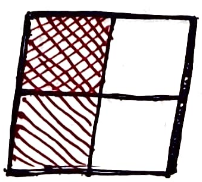
Actual Value

Prediction	00 P	01 N
P	TP	FP
N	FN	TN
	10	11

$$\text{precision} = \frac{TP}{TP+FP}$$

FP ↑ Precision ↓

$$\text{Recall} = \frac{TP}{TP+FN}$$



FN ↑ Recall ↓

Precision ⟶ FP.
Recall ⟶ FN

L08 - Classification Evaluation part - 2

* When to focus on precision and when on recall?

(Medical Test) Cancer → test result → Negative. But actually $\boxed{\text{positive}}$

↳ In this ~~cabl~~ case, false negative is dangerous

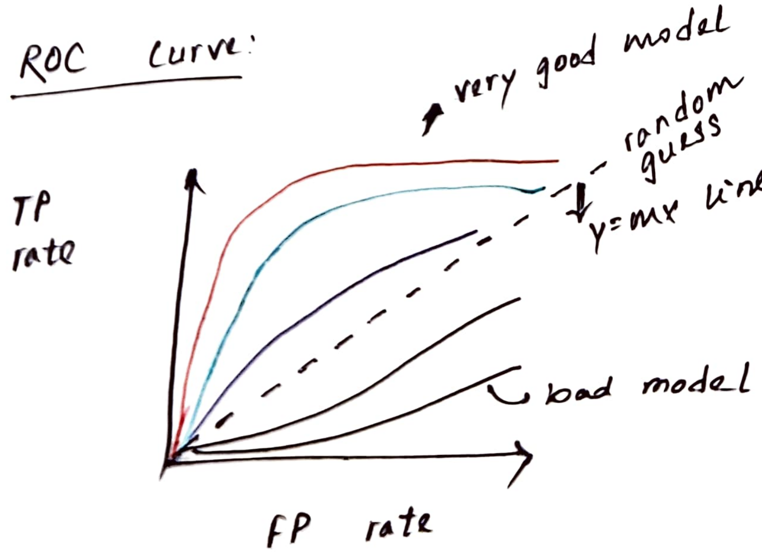
(Spam Mail) ⇒ predicting an important mail as spam is wrong

$$\boxed{F1 \ score = \frac{2 \times P \times R}{P + R}}$$

P = Precision
R = Recall

⇓ Combination of F1

ROC curve:



• Above the "random guess" line, good model.

• If a curve goes below the "random guess" line, it will be a bad model. The lower it goes, the worse it becomes.

AUC (Area Under the ~~the~~ Curve):

AUC ↑ model quality ↑

AUC ↓ model quality ↓

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

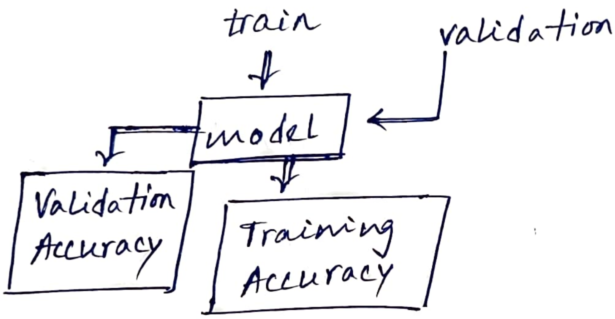
$$= \frac{TP + TN}{N}$$

N= Number of samples.

L 09 - Bias Variance - Overfitting and Underfitting

Dataset
• 3 partitions

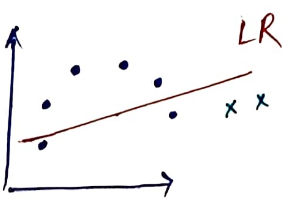
→ Train
→ Validation — used while training 80%
→ Test — 20%



• test dataset will be hidden

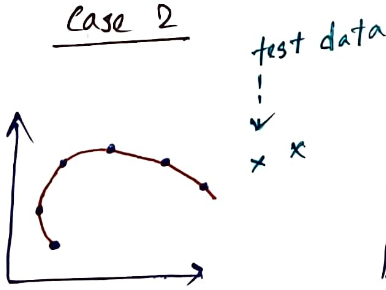
$\boxed{\text{test accuracy}}$

case 1

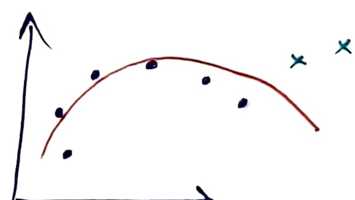


train - Low Accuracy
test - Low Accuracy

case 2



High Accuracy
Low Accuracy



Good Accuracy
Good Accuracy
⇓
Perfect fit
• Low bias
• Low varias.

∴ Bias → Opposite of Accuracy
Variance → " " Consistency.

* ↓ ~~Bo~~ Bias ↓ Variance → Good Model.

⇒ case 2 has high variance, because ~~inacc~~ inconsistent.
⇒ case 1 has high bias, because inaccurate.

→ overfit
→ underfit.

17

Train Acc →	90%	80%	30%	90% *	85%
Test Acc →	50%	76%	28%	82%	83%
	overfit	good fit	underfit	good / overfit	perfect

- in this case, try to get another unseen dataset to check.

$\boxed{\begin{array}{l} 95\% \\ 92\% \end{array}}$

Amazing!

- but hard to find
- depends on dataset

if F1 > 80% → good model

L 10 - Cross Validation

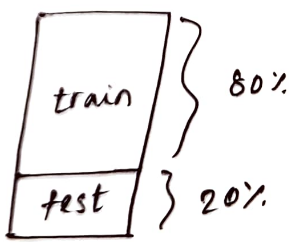
* if underfit happens
  - 1. increase training time.
  - 2. increase features.

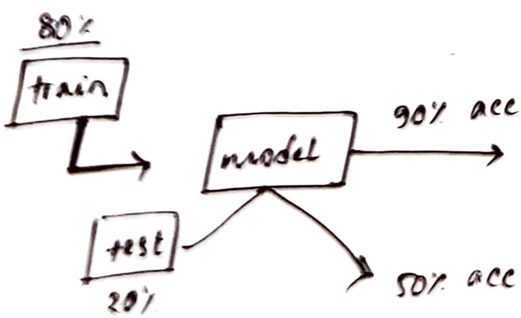
* overfitting
  1. cross validation.
  2. Regularization

To handle overfit, we have regularization methods.

Lasso	Ridge	Elastic Net

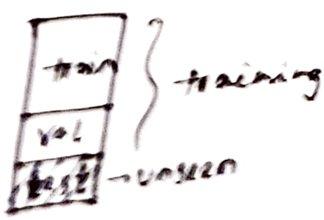
Dataset

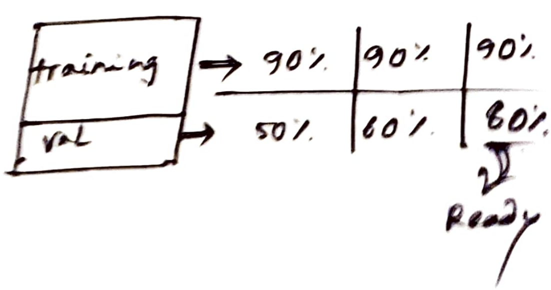




Cross Validation (K-fold):

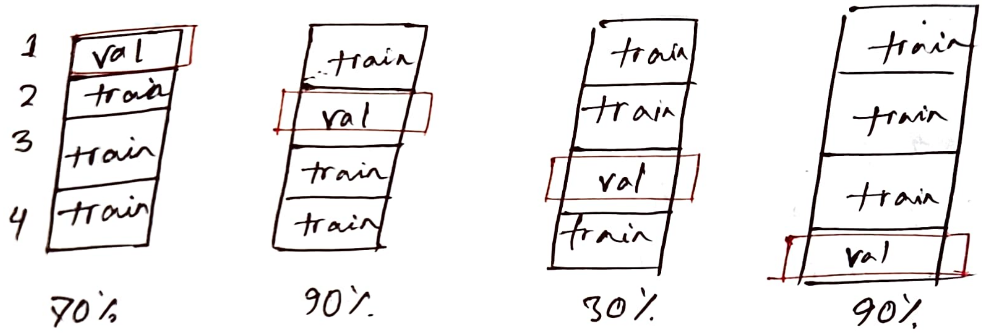
K=4





19

$K=4$



$$\frac{70\% + 90\% + 30\% + 90\%}{4} = 70\% \text{ (Avg)}$$

Validation Accuracy