

G.H. RAISONI COLLEGE OF ENGINEERING

2020-2021 EVENTERM

CAE - I EXAMINATION SUMMER - 2021 (ONLINE MODE)

DEPARTMENT : ARTIFICIAL INTELLIGENCE

SEMESTER / SECTION : 4th / A

DATE : 12/02/2021

SUBJECT : MACHINE LEARNING ALGORITHMS

ROLL NO : A - 58

NAME : SHIVAM TAWARI

REG. NO. : 2019AAIE1117628

COL.

Obs.	Temp(x)°c	Humidity(y)%	Rain(Condition)	Dist
1	27.8	76	Yes	2.69
2	28.2	76	Yes	2.44
3	28.7	80	No	2.19
4	28.6	81.6	Yes	3.73
5	27.7	89.4	Yes	11.55
6	30.5	89.9	No	11.93
7	26.87	81.4	Yes	4.46
8	25.9	85	No	7.91
9	36	90	No	13.60
10	31.8	88	Yes	10.24
11	35.7	70	No	10.06

Shivam Tawari

Pg. no ①

Distance metric: Euclidean Distance

Observation 1:

$$d = \sqrt{(29.6 - 27.8)^2 + (78 - 76)^2} = 2.69$$

Observation 2:

$$d = \sqrt{(28.2 - 29.6)^2 + (78 - 76)^2} = 2.44$$

Observation 3:

$$d = \sqrt{(29.6 - 28.7)^2 + (78 - 80)^2} = 2.19$$

Observation 4:

$$d = \sqrt{(29.6 - 28.6)^2 + (78 - 81.6)^2} = 3.73$$

Observation 5:

$$d = \sqrt{(29.6 - 27.7)^2 + (78 - 89.4)^2} = 11.55$$

Observation 6:

$$d = \sqrt{(29.6 - 30.5)^2 + (78 - 89.9)^2} = 11.93$$

Observation 7:

$$d = \sqrt{(29.6 - 26.7)^2 + (78 - 81.4)^2} = 4.46$$

Observation 8:

$$d = \sqrt{(29.6 - 25.9)^2 + (78 - 85)^2} = 7.91$$

~~Hawaii~~ Pg. no. 2

Observation 9:

$$d = \sqrt{(29.6 - 36)^2 + (78 - 90)^2} = 13.60$$

Observation 10:

$$d = \sqrt{(31.8 - 29.6)^2 + (88 - 78)^2} = 10.24$$

Observation 11:

$$d = \sqrt{(29.6 - 35.7)^2 + (78 - 70)^2} = 10.06$$

Min. 3 distances are 2.19, 2.44, 2.69,
and their conditions are No, Yes,
Yes respectively.

Majority is Yes.

Therefore when temp: 29.6°C and
humidity: 78% ~~it's~~ it'll probably
rain.

Q1.

b.

Over-fitting :

- Over-fitting refers to a model that models the training data too well.
- Over-fitting model refers to having a high variance.
- Over-fitting models are more nonparametric.
- Over-fitting are non-linear models.

Ex. Decision trees are a nonparametric machine learning algorithm that is very flexible and subject to overfitting.

Under-fitting :

- Under-fitting refers to a model that can neither model the training data nor generalize to new data.
- Has a very low variance.
- Under-fitting model has high bias.
- Underfit model does not gives a good metric performance.

Ex. ~~Progs~~

Underfitting happens when we are trying to fit non-linear data into the linear regression.

CO2.

b.

Supervised Learning

① Uses known and labeled data as input.

② Very complex computation.

③ Uses off-line analysis

④ Number of classes are known.

⑤ Accurate and reliable results.

Unsupervised Learning

Uses unknown data as input

Less computational complexity.

Uses real time analysis of data.

Numbers of classes are not known.

Moderate accurate and reliable results.