

AML ASSIGNMENT

1. What is semi supervised machine learning? Explain with example?

Semi-supervised machine learning is a combination of **supervised** and **unsupervised machine learning** methods. In semi supervised learning an algorithm learns from a dataset that includes both **labelled** and **unlabelled data**, usually mostly unlabeled.

A common example of an application of semi supervised learning is a text document classifier. This is the type of situation where semi-supervised learning is ideal because it would be nearly impossible to find a large amount of labeled text documents. This is simply because it is not time efficient to have a person read through entire text documents just to assign it a simple classification. So semi-supervised learning allows the algorithms to learn from a small amount of labelled text documents while still classifying a large amount of unlabelled text documents in the training data.

How semi supervised learning works:-

1. Train the model with the small amount of labelled training data until it gives a good result.
2. Then use it with the unlabelled training dataset to predict the outputs which are pseudo labels since they may not be quite accurate.
3. Link the labels from the labelled training data with the pseudo labels created in the previous step.
4. Link the data inputs in the labelled training data with the inputs in the unlabelled data.
5. Then train the model the same way as you did with the labelled set in the beginning in order to decrease the error and improve the models accuracy.

2. How will you decide the k -value in k -NN algorithm?

- * Try with different values and choose the best one.
- * k -value must always be odd.

3. How does the efficiency and accuracy of KNN search change as k increases?

- * If we have a large number of training set the accuracy should increase.
- * The larger the training set less the efficiency.
- * The time to calculate the prediction will increase as computational complexity increases.

4. Why is KNN a lazy learning algorithm?

- * No learning of the model/algorithm.
- * It memorizes the training set.

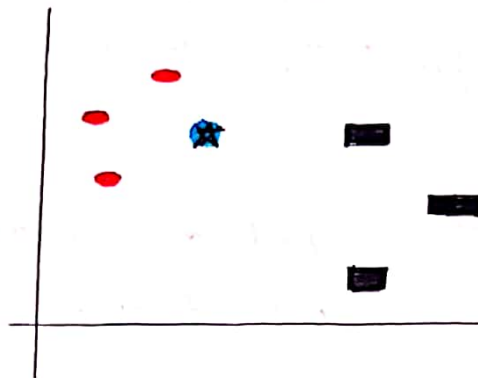
5. Why is KNN a non-parametric algorithm?

- * Because it makes no assumptions about the functional form of problem being solved.

6. When do we use KNN algorithm?

- * It is used for both classification and regression problems.
- * Widely used for classification problems in industry.
- * Used for its easy interpretation and low calculation time.
- * Hence its predictive power increases.

7. How does the KNN algorithm work to classify the blue star?



- * First we need to consider the **K-value**.
- * Then using the euclidean distance formula, the distance from the query point to other points will be calculated.
- * The **K-nearest** points will be considered for classification.
- * The blue star will be classified according to the most frequently occurring points.

8. Assume a boolean target function and a 2-D instance space. Determine how the knearest neighbour learning algorithm would classify the new instance x_q for $k=1, 3, 5$ and 7 . The $+$ and $-$ signs in the instance space refer to the positive and negative examples respectively.

Distance from query instance	classification
1.00	+
1.35	-
1.40	-
1.60	-
1.90	+
2.00	+
2.20	-
2.40	+
2.80	-

In the case 1 when $k=1$

The query point x_q will be classified as a positive example because the nearest example to x_q is positive.

when $k=3$

The query point x_q will be classified as a negative example because the negative examples occur frequently when $k=3$.

when $k=5$

The query point x_q will be classified as a negative example because the negative examples are more frequent when $k=5$.

when $k=7$

The query point x_q will be classified as negative example because the negative examples are more frequently occurring when $k=7$. i.e. there are 3 positive examples and 4 negative examples. So the negative examples are more frequent.