

Worksheet 7

Machine learning

1) A

2) A

3) B

4) C

5) A

6) C

7) B

8) B

9)

10)

Random forest algorithm avoids and prevents overfitting by using multiple trees. The results are not accurate. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy

11)

Scaling is required to rescale the data and it's used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better. scaling is the process of normalising the range of features in a dataset. Real-world datasets often contain features that are varying in degrees of magnitude, range and units. Therefore, in order for machine learning models to interpret these features on the same scale, we need to perform feature scaling.

Scaling is a common dental procedure for patients with gum disease. This is a type of dental cleaning that reaches below the gumline to remove plaque buildup. The process of scaling and root planing the teeth is often referred to as a deep cleaning.

The first is standard scaling (or z-scaling) and is calculated by subtracting the mean and dividing by the standard deviation. The second is min-max scaling and is calculated by subtracting by the minimum value and dividing by the difference between the maximum and minimum values.

12)

The main advantages:

We can use fixed learning rate during training without worrying about learning rate decay.

It has straight trajectory towards the minimum and it is guaranteed to converge in theory to the global minimum if the loss function is convex and to a local minimum if the loss function is not convex.

It has unbiased estimate of gradients. The more the examples, the lower the standard error.

13)

A classifier is only as good as the metric used to evaluate it.

If you choose the wrong metric to evaluate your models, you are likely to choose a poor model, or in the worst case, be misled about the expected performance of your model.

Choosing an appropriate metric is challenging generally in applied machine learning, but is particularly difficult for imbalanced classification problems. Firstly, because most of the standard metrics that are widely used assume a balanced class distribution, and because typically not all classes, and therefore, not all prediction errors, are equal for imbalanced classification.

14)

The F-score (also known as the F1 score or F-measure) is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score. F-measure formula: $F\text{-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$

15)

The `fit(data)` method is used to compute the mean and std dev for a given feature to be used further for scaling. The `transform(data)` method is used to perform scaling using mean and std dev calculated using the `. fit()` method. The `fit_transform()` method does both fits and transform.

STATISTICS

- 1) B
- 2) D
- 3) C
- 4) B
- 5) C
- 6) A
- 7) C
- 8) B
- 9) A
- 10) A
- 11) C
- 12) A
- 13) D
- 14) D
- 15) A