

Machine Learning
Worksheet - 4

Latika Taneja

1. > (c) between -1 & 1
2. > (c) Recursive feature elimination
3. > (c) hyperplane
4. > (A) Logistic Regression
5. > (c) Old coefficient of 'x' $\div 2.205$
6. > (B) increases
7. > (c) Random forest are easy to interpret.
- (8.) (B) Principle components are calculated using unsupervised learning techniques
(A) (c) Principle components are linear combination of linear variables

(9) (B), (C)

(10) (A), (B), (C)

(11) 12th question (Interquartile range) from statistics worksheet before

(12) Bagging.

- Simplest way of combining predictions that belong to same type
- It tries to solve overfitting problem

Boosting

- A way of combining prediction that belong to the different types.
- It tries to reduce bias

(13)

Adjusted R-Squared

- It measures the proportion of variation explained by only those independent variables that really help in explaining the dependent variables.

Formula:

$$R^2_{\text{adjusted}} = \frac{1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}}$$

where:

R^2 = sample R-square

p = number of predictors

n = total sample size.

(14)

Normalization

It is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1.

Formula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

$$x_{\max} - x_{\min}$$

Standardization

- It is a scaling technique where the values are centered around the mean with a unit standard deviation.
- This means that the mean of the attribute becomes 0 and the resultant distribution has a unit standard deviation.

• Formula:

$$x' = \frac{x - \mu}{\sigma}$$

μ : mean

σ : standard deviation

(15) • Cross validation is a statistical method or a resampling procedure used to evaluate the skill of machine learning models on a limited data sample.

• Advantage:

It reduces over-fitting by splitting the dataset into multiple folds and train the algorithm on different folds.

- Disadvantage.

It increases training time. Earlier you had to train your model only on one training set but with ~~the~~ cross validation you have to train your model on multiple training set.