

# Worksheet\_set\_1

## Machine Learning Assignment

Q.1 Answer: A

Q.2 Answer: A

Q.3 Answer: A

Q.4 Answer: B

Q.5 Answer: C

Q.6 Answer: B

Q.7 Answer: D

Q.8 Answer: D

Q.9 Answer: A

Q.10 Answer: A

Q.11 Answer: D

Q.12 Answer: A, B, C

### Q.13 Answer:

When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort this overfitting problem by restricting the degrees of freedom of a given equation i.e. simply reducing the number of degrees of a polynomial function by reducing their corresponding weights. In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable(Y). So, regularization constraints the weights of such features to avoid overfitting.

Regularization helps to reduce the variance of the model, without a substantial increase in the bias. If there is variance in the model that means that the model won't fit well for dataset different than training data. The tuning parameter  $\lambda$  controls this bias and variance tradeoff. When the value of  $\lambda$  is increased up to a certain limit, it reduces the variance without losing any important properties in the data. But after a certain limit, the model will start losing some important properties which will increase the bias in the data. Thus, the selection of good value of  $\lambda$  is the key. The value of  $\lambda$  is selected using cross-validation methods. A set of  $\lambda$  is selected and cross-validation error is calculated for each value of  $\lambda$  and that value of  $\lambda$  is selected for which the cross-validation error is minimum.

# Worksheet\_set\_1

## Q.14 Answer:

To regularize the model, a Shrinkage penalty is added to the cost function. Let's see different types of regularizations in regression:

LASSO  
RIDGE

### **LASSO (Least Absolute Shrinkage and Selection Operator) Regression (L1 Form)**

LASSO regression penalizes the model based on the sum of magnitude of the coefficients. The regularization term is given by  $\text{regularization} = \lambda * \sum ||\beta_j||$  Where,  $\lambda$  is the shrinkage factor.

### **Ridge Regression (L2 Form)**

Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients. The regularization term is given by  $\text{regularization} = \lambda * \sum ||\beta_j||^2$  Where,  $\lambda$  is the shrinkage factor.

### ***Difference between Ridge and Lasso*** $\lambda * \sum ||\beta_j||$ $\lambda * \sum ||\beta_j||^2$

Ridge regression shrinks the coefficients for those predictors which contribute very less in the model but have huge weights, very close to zero. But it never makes them exactly zero. Thus, the final model will still contain all those predictors, though with less weights. This doesn't help in interpreting the model very well. This is where Lasso regression differs with Ridge regression. In Lasso, the L1 penalty does reduce some coefficients exactly to zero when we use a sufficiently large tuning parameter  $\lambda$ . So, in addition to regularizing, lasso also performs feature selection

## Q.15 Answer:

The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors, and omitted variables.

# **Worksheet\_set\_1**

## **Python Worksheet Solution**

Q.1 Answer: C

Q.2 Answer: A

Q.3 Answer: C

Q.4 Answer: A

Q.5 Answer: D

Q.6 Answer: C

Q.7 Answer: A

Q.8 Answer: A

Q.9 Answer: A & C

Q.10 Answer: A & B

For Q.11 to 15 Python Notebook is uploaded in Github

# Worksheet\_set\_1

## Statistic Worksheet Solution

Q.1 Answer: A

Q.2 Answer: A

Q.3 Answer: B

Q.4 Answer: D

Q.5 Answer: C

Q.6 Answer: B

Q.7 Answer: B

Q.8 Answer: A

Q.9 Answer: C

### Q.10 Answer:

Normal Distribution means uniform spread of data across mean.

If something is said to follow the normal distribution, it means in the simplest terms that most of the data lies around the average. An easy example is the distribution of test grades in schools. Most people will score around the average, with a few high scores and a few low scores. This means that most people get C's, while only a few get A's and F's.

### Q.11 Answer

Handling missing values is the important step while building your model. It will impact the result if not handled well. The missing values occur in data due to many reasons, such as problems occurred during extraction or data collection process.

- **Imputing**

- Continuous features:

- Most common imputed value is generally the mean but be vary. You don't want information leakage in your train and test set. It's a general approach to

## Worksheet\_set\_1

find the mean of training data and impute both your training and test data with that value.

- Median is a good approach as well, try making a histogram and see the distribution of your data and from there you can decide what would make more sense, mean or median
- Categorical features
  - For categorical variables, you can use forward fill or backward fill if you don't really want to get too much into the weeds and still want the null values to be randomly filled.
  - We can also use the most common value of the column to fill up the null values. Do check the distribution before doing that.

In the end, it depends on your problem and data. Depending on the techniques, we can play with them and see how using different methods results can be changed

### Q.12 Answer

A/B testing a testing method that is widely used by marketers for conversion rate optimization. They evaluate two different alternatives of a product, service, landing page, or process by splitting traffic into two equal sizes. The main purpose of A/B testing is to understand the audience better so that you can choose the version that works better.

Businesses most of the time want to compare multiple combinations (e.g. price points, colors etc.) and A/B/C/D/... testing approach starts to be difficult to implement, especially for low-traffic websites. For such cases, approaches like multi-armed bandit can be preferred.

### Q.13 Answer

Not really.

Although it's a popular solution to missing data, despite its drawbacks. Mainly because it's easy. It can be really painful to lose a large part of the sample you so carefully collected.

### Q.14 Answer

A Simple Linear Regression allows you to determine functional dependency between two sets of numbers (Continuous values). For example, we can use regression to determine the relation between ice cream sales and average temperature.

Since we are talking about functional dependency between two sets of variables, we need an independent variable and one dependent variable. In the example above, if change in temperature

## Worksheet\_set\_1

leads to change in ice cream sales then, temperature is independent variable and sales is dependent variable.

Dependent variable is also called as criterion, response variable or label. It is denoted by Y.

The independent variable is also referred as covariates, predictor or features. It is denoted by X.

### **Q.15 Answer:**

Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are **two main branches** of statistics

- Inferential Statistic.
- Descriptive Statistic.

### **Inferential Statistics:**

Inferential statistics used to make inference and describe about the population. These stats are more useful when it's not easy or possible to examine each member of the population.

### **Descriptive Statistics:**

Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.