MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

   Answer: A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

   Answer: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is ______?

   Answer: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

   Answer: A) Regression

5. Which of the following is the reason for over fitting condition?

   Answer: C) Low bias and high variance

6. If output involves label then that model is called as:

   Answer: B) Predictive modal

7. Lasso and Ridge regression techniques belong to _________?

   Answer: A) Cross validation

8. To overcome with imbalance dataset which technique can be used?

   Answer: A) Cross validation

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

   Answer: A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

    Answer: A) True

11. Pick the feature extraction from below:

    Answer: B) Apply PCA to project high dimensional data

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

    Answer: A) We don’t have to choose the learning rate.

    B) It becomes slow when number of features is very large.

13). Explain the term regularization

Answer:

Regularization in Machine Learning

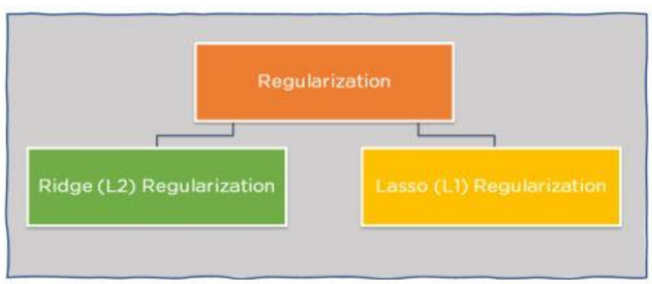
Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Regularization on an over-fitted model

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Regularization Techniques :

There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.



Regularization significantly reduces the variance of the model, without substantial increase in its bias. *Till* a point, this increase in λ is beneficial as it is only reducing the variance (hence avoiding overfitting), without losing any important properties in the *data* A simple relation for linear regression looks like this. Here Y represents the learned relation and *β represents the coefficient estimates for different variables or predictors(X).*

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

Develop Machine Learning Applications for Business : The loss function called 'the residual sum of square' is mostly used for linear regression. Here's what it looks like:

$$\text{RSS} = \sum_{j=1}^{m} \left(Yi - Wo - \sum_{i=1}^{n} Wi\, Xji \right)^2$$

Regularization Using Python in Machine Learning:

We start by importing all the necessary modules.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

We then load the Boston Housing Dataset from sklearn's datasets.

```
# Loading pre-defined Boston Dataset
boston_dataset = datasets.load_boston()
```

We then split our data into training and testing sets.

```
x_train, x_test, y_train, y_test = train_test_split(
boston_pd.iloc[:, :-1], boston_pd.iloc[:, -1],
test_size = 0.25)

print("Train data shape of X = % s and Y = % s : "%(
x_train.shape, y_train.shape))

print("Test data shape of X = % s and Y = % s : "%(
x_test.shape, y_test.shape))
```

```
Train data shape of X = (379, 13) and Y = (379,) :
Test data shape of X = (127, 13) and Y = (127,) :
```

14). Which particular algorithm are used for regularization?

Answer:

Algorithm used for Regularization:

The term ‘regularization’ refers to a set of techniques that regularizes learning from particular features for traditional algorithms or neurons in the case of neural network algorithms. It normalizes and moderates’ weights attached to a feature or a neuron so that algorithms do not rely on just a few features or neurons to predict the result. This technique helps to avoid the problem of overfitting.To understand regularization, let’s consider a simple case of linear regression. Mathematically, linear regression is stated as below:

$$y = w_0 + w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n$$

There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)
2. Lasso (L1 Norm)
3. Dropout

Ridge and Lasso can be used for any algorithms involving weight parameters, including neural nets. Dropout is primarily used in any kind of neural networks e.g. ANN, DNN, CNN or RNN to moderate the learning. Let’s take a closer look at each of the techniques.

Ridge Regression (L2 Regularization): Ridge regression is also called L2 norm or regularization When using this technique, we add the sum of weight’s square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda \sum_{i=1}^{n} Wi^2$$

As seen above, the original loss function is modified by adding normalized weights. Here normalized weights are in the form of squares.

You may have noticed parameters λ along with normalized weights. λ is the parameter that needs to be tuned using a cross-validation dataset. When you use λ=0, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimize weight's square and will end up taking the parameters' value as zero.

Now the parameters are learned using a modified loss function. To minimize the above function, parameters need to be as small as possible. Thus, L2 norm prevents weights from rising too high.

Lasso Regression (L1 Regularization):

Also called lasso regression and denoted as below:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^2 + \lambda\sum_{i=1}^{n}|Wi|$$

This technique is different from ridge regression as it uses absolute weight values for normalization. λ is again a tuning parameter and behaves in the same as it does when using ridge regression.

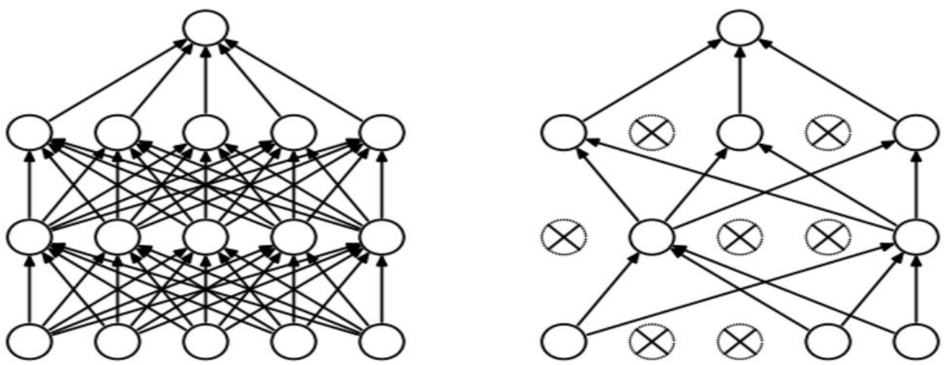
As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.

Dropout:

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

In neural nets, fully connected layers are more prone to overfit on training data. Using dropout, you can drop connections with *1-p* probability for each of the specified layers. Where *p* is called keep probability parameter and which needs to be tuned.



(a) Standard Neural Net (b) After applying dropout.

With dropout, you are left with a reduced network as dropped out neurons are left out during that training iteration.

Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize better on unseen data. However, it is important to note that Dropout takes more epochs to train compared to training without Dropout (If you have 10000 observations in your training data, then using 10000 examples for training is considered as 1 epoch).

Along with Dropout, neural networks can be regularized also using L1 and L2 norms. Apart from that, if you are working on an image dataset, image augmentation can also be used as a regularization method.

For real-world applications, it is a must that a model performs well on unseen data. The techniques we discussed can help you make your model learn rather than just memorize.

15). Explain the term error present in linear regression model?

Answer:
Linear regression: most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.

The next row in the 'Coefficients' table is income. This is the row that describes the estimated effect of income on reported happiness:

The Estimate column is the estimated **effect**, also called the **regression coefficient** or r^2 value. The number in the table (0.713) tells us that for every one unit increase in income (where one unit of income = 10,000) there is a corresponding 0.71-unit increase in reported happiness (where happiness is a scale of 1 to 10).

The Std. Error column displays the **standard error** of the estimate. This number shows how much variation there is in our estimate of the relationship between income and happiness

The t value column displays the **test statistic**. Unless you specify otherwise, the test statistic used in linear regression is the *t*-value from a two-sided t-test. The larger the test statistic, the less likely it is that our results occurred by chance.

The Pr(>| t |) column shows the ***p*-value**. This number tells us how likely we are to see the estimated effect of income on happiness if the null hypothesis of no effect were true.

Because the *p*-value is so low ($p < 0.001$), we can **reject the null hypothesis** and conclude that income has a statistically significant effect on happiness. The last three lines of the model summary are statistics about the model as a whole. The most important thing to notice here is the *p*-value of the model. Here it is significant ($p < 0.001$), which means that this model is a good fit for the observed data.