# MACHINE LEARNING

only one option is correct, choose the Which of the following methods do we use to find the best correct option

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

**OPTION - A) Least Square Error**

1. **Maximum Likelihood**
2. **Logarithmic Loss**
3. **Both A and B**

**ANSWER- OPTION-B Maximum likelihood**

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

1. Linear regression is sensitive to outliers
2. linear regression is not sensitive to outliers
3. Can’t say
4. none of these

Answer= **A linear regression is sensitive to outliers**

# A line falls from left to right if a slope is ?

1. **Positive**
2. **Negative**
3. **Zero**
4. **Undefined**

ANSWER=b- negative

# 4- Which of the following will have symmetric relation between dependent variable and independent variable**?**

1. **Regression**
2. **Correlation**
3. **Both of them**
4. **None of these**

**Answer= b- correlation**

# Which of the following is the reason for over fitting condition?

A) High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none for over fitting condition

Answer=b-lowbias and low high variance

# If output involves label then that model is called as:

**A) Descriptive model B) Predictive modal**

**C) Reinforcement learning D) All of the above**

Answer=b- predicitive model

# Lasso and Ridge regression techniques

belong to \_\_\_\_ \_ ?

**A) Cross validation B) Removing outliers**

**C) SMOTE D) Regularization**

Answer = **D) Regularization**

# To overcome with imbalance dataset which technique can be used?

**A) Cross validation B) Regularization**

**C) Kernel D) SMOTE**

Answer=d smote

# The AUC Receiver Operator Characteristic (AUCROC) curve is an

evaluation metric for binary classification problems. It uses make graph?

# \_\_\_\_\_ to

**A) TPR and FPR B) Sensitivity and precision**

**C) Sensitivity and Specificity D) Recall and precision**

Answer=A-trp and fpr

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

**A) True B) False**

Answer=A-true

# 1. Pick the feature extraction from below:

**A) Construction bag of words from a email B) Apply PCA to project high dimensional data C) Removing stop words D) Forward selection**

Answer=B-apply pca to project high dimensional data

# In Q12, more than one options are correct, choose all the correct options:

1. Which of the following is true about

# Normal Equation used to compute the coefficient of the Linear Regression**?**

**A) We don’t have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate. D) It does not make use of dependent variable.**

Answer= b **. B) It becomes slow when number of features is very large.**

1. Explain the term regularization? Answer=Regularization is a technique used in various fields, particularly in machine learning and statistics, to prevent overfitting and improve the generalization ability of models. Overfitting occurs when a model learns to perform very well on the training data but fails to generalize effectively to new, unseen data. Regularization methods aim to address this issue by adding additional constraints or penalties to the model's optimization process, which encourages the model to be simpler and less prone to fitting noise in the training data.

The primary goal of regularization is to find a balance between fitting the training data well (low training error) and having good generalization performance on new, unseen data (low test error).

Regularization techniques work by introducing a penalty term to the objective function that the model is trying to optimize. This penalty discourages the model from assigning excessive importance to certain features or from learning complex relationships that might be driven by noise.

There are different types of regularization techniques, including:

L1 Regularization (Lasso): This method adds a penalty proportional to the absolute values of the model's coefficients. It encourages the model to drive some coefficients to exactly zero, effectively performing feature selection. L1 regularization can lead to sparse models where only a subset of features are important.

L2 Regularization (Ridge): L2 regularization adds a penalty proportional to the square of the model's coefficients. It discourages large coefficient values and promotes a more evendistribution of feature importance. L2 regularization is less likely to result in exactly zero coefficients, but it can still lead to reduced magnitudes.

Elastic Net Regularization: Elastic Net combines both L1 and L2 regularization, allowing for a balance between feature selection (L1) and coefficient shrinkage (L2). It aims to capture the benefits of both L1 and L2 regularization while mitigating their individual limitations.

Dropout: A regularization technique often used in neural networks, dropout involves randomly setting a fraction of the neurons' outputs to zero during each training iteration. This helps prevent complex co-

adaptations of neurons, making the network more robust and less likely to overfit.

Early Stopping: While not a penalty-based technique, early stopping is a form of regularization that monitors the model's performance on a validation set and stops training when the performance starts deteriorating. This prevents the model from continuing to optimize and potentially overfitting the training data.

Data Augmentation: Another form of regularization involves artificially increasing the size of the training dataset by applying various transformations to the existing data, such as rotations, translations, or other modifications. This helps expose the model to a wider range of variations in the data distribution. Regularization plays a crucial role in achieving better generalization and robustness in machine learning models. By adding these constraints or penalties, regularization methods encourage models to focus on capturing the most important underlying patterns in the data while avoiding overfitting noise or irrelevant details present in the training data

15. Explain the term error present in linear regression equation?

Answer=In the context of a linear regression equation, the term "error" refers to the difference between the actual observed values of the dependent variable and the values predicted by

the linear regression model. This error is also commonly known as the "residual."

Mathematically, the linear regression equation can be represented as:

=

0

+

1

1

+

2

2

+

…

+

+

y=β 0

+β 1

x 1

+β 2

x 2

+…+β n

x n

+ε

Where:

y is the dependent variable (the variable we are trying to predict).

1

,

2

,

…

,

x 1

,x 2

,…,x

n

are the independent variables (features) that are used to make predictions.

0

,

1

,

…

,

β 0

,β 1

,…,β n

are the coefficients (parameters) that the model aims to estimate to fit the data.

ε represents the error term, which is the difference between the actual observed

y values and the predicted values

^ y

^

.

In essence, the error term captures the variability in the data that the linear regression model cannot explain. It represents

the noise or randomness in the relationship between the dependent variable and the independent variables. The goal of linear regression is to minimize the sum of squared errors (SSE) or residuals, which represents the sum of the squared differences between the actual values and the predicted values:

SSE

=

∑

= 1

(

−

^

) 2

SSE=∑

i=1 n

(y i

−

y

^

i

)

2

Minimizing the SSE results in finding the coefficients

0

,

1

,

…

,

β 0

,β

1

,…,β n

that best fit the data and provide the line that best represents the relationship between the variables.

The error term is an important concept in linear regression because it helps us understand the accuracy and reliability of our model's predictions. Larger errors indicate that the model is not capturing the underlying patterns in the data well, while smaller errors indicate a better fit. However, it's important to note that some level of error is inevitable due to the inherent variability and noise in real-world data. The goal of linear regression is to find a balance between minimizing the error an? d avoiding overfitting the model to the noise in the data.

14. Which particular algorithms are used for regularization?

regularization in machine learning, refers to a set of techniques that help the machine to learn more than just memorize. Before we explore the concept of regularization in detail, let’s

discuss what the terms ‘learning’ and ‘memorizing’ mean from the perspective of machine learning.

When you train a machine learning model, and it is able to deliver accurate results on training data, but provides relatively poor results on unseen data or test dataset, you can say your model is memorizing more than generalizing.

Let’s say you are doing a cat vs dog classification and your trained model delivers 96% accuracy on training data, but when you run the same model using test dataset you get 85% accuracy on that data set – your model is memorizing more than generalizing

Let’s take a real-world scenario. Assume, an e-commerce company wishes to build a model to predict if the user would buy a product or not, given his/her usage history for last 7 days, and use the data for better decision making for retargeted digital advertisements. The usage history may include the number of pages visited, total time spent, the number of searches done, average time spent, page revisits, and more.

They build a model and it delivers accurate results on existing data but when they try to predict with unseen data, it doesn’t deliver a good result. Here, it can be concluded that the model does more of memorization than learning.

So what’s really happening with the above examples? One strong possibility is that the model has an overfitting problem, which results in relatively poor performance on unseen data. So essentially, memorizing is taking place here and not learning.

In general, if your model has a significant difference between evaluation metrics for training dataset and testing dataset, then it is said to have an overfitting problem.

Now that we have explored the concepts of memorizing, learning, and overfitting, let’s move to regularization and regularization techniques.

**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 = w0 + w1x1 + w2x2 + ….. + wnxn***

where y is the value to be predicted;

x1, x2, …., xn are features that decides the value of y;

w0 is the bias;

w1, w2, ….., wn are the weights attached to x1, x2, …., xn relatively.

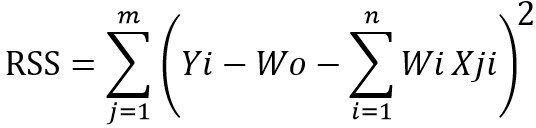
Now to build a model that accurately predicts the *y* value, we need to optimize above mentioned bias and weights.

To do so, we need to use a loss function and find optimized parameters using gradient descent algorithms and its variants.

To know more about building a machine learning application and the process, check out below blog:

How to 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 :



Next, we will learn bias (or intercept) and weights (also identified as parameters and coefficients) using the optimization algorithm (gradient descent) and data. If your dataset does have noise in it, it will face overfitting problem and learned parameters will not generalize well on unseen data.

To avoid this, you will need to regularize or normalize your weights for better learning.

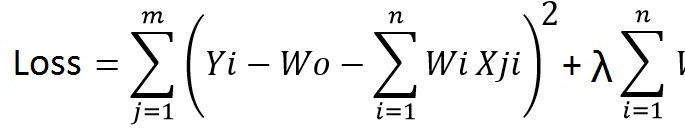
There are three main regularization techniques, namely:

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.

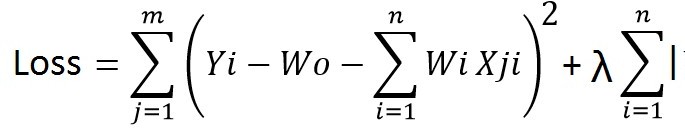
When using this technique, we add the sum of weight’s square to a loss function and thus create a new loss function which is denoted thus:

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:

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.

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

Be it an over-fitting or under-fitting problem, it will lower down the overall performance of a machine learning model. To get the best out of machine learning models, you must optimize and tune them well. At eInfochips, we deliver machine learning services that help businesses optimize the utilization AI technology. We have machine learning capabilities across cloud, hardware, neural networks, and open source frameworks. Connect with our team to learn more about how machine learning can be useful