ASSIGNMENT 5:

Answer 1 : A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data.

Training a model includes following steps :

Step 1: Begin with existing data. Machine learning requires us to have existing data—not the data our application will use when we run it, but data to learn from. ...

Step 2: Analyse data to identify patterns. ...

Step 3: Make predictions.

Answer 2:

The theory asserts that when the performance of all optimization methods is averaged across all conceivable problems, they all perform equally well. It indicates that no one optimum optimization algorithm exists. Because of the strong link between optimization, search, and machine learning, there is no one optimum machine learning method for predictive modelling tasks like classification and regression.

Answer 3:

K-fold cross-validation approach divides the input dataset into K groups of samples of equal sizes. These samples are called folds. For each learning set, the prediction function uses k-1 folds, and the rest of the folds are used for the test set.

Answer 4:

The bootstrap method is a statistical technique for estimating quantities about a population by averaging estimates from multiple small data samples.

Importantly, samples are constructed by drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen.

Answer 5 :

Cohen’s kappa is a metric often used to assess the agreement between two raters. It can also be used to assess the performance of a classification model.

For example, if we had two bankers and we asked both to classify 100 customers in two classes for credit rating (i.e., good and bad) based on their creditworthiness, we could then measure the level of their agreement through Cohen’s kappa.

Similarly, in the context of a classification model, we could use Cohen’s kappa to compare the machine learning model predictions with the manually established credit ratings.

Like many other evaluation metrics, Cohen’s kappa is calculated based on the confusion matrix. However, in contrast to calculating overall accuracy, Cohen’s kappa takes imbalance in class distribution into account and can, therefore, be more complex to interpret.

Answer 6: Ensemble modeling is a process where multiple diverse models are created to predict an outcome, either by using many different modeling algorithms or using different training data sets. The ensemble model then aggregates the prediction of each base model and results in once final prediction for the unseen data.

Answer 7: Descriptive modeling is a mathematical process that describes real-world events and the relationships between factors responsible for them. The process is used by consumer-driven organizations to help them target their marketing and advertising efforts.

In descriptive modeling, customer groups are clustered according to demographics, purchasing behavior, expressed interests and other descriptive factors. Statistics can identify where the customer groups share similarities and where they differ. The most active customers get special attention because they offer the greatest [ROI](https://www.techtarget.com/searchcio/definition/ROI) (return on investment).

The main aspects of descriptive modeling include:

* Customer segmentation: Partitions a customer base into groups with various impacts on marketing and service.
* Value-based segmentation: Identifies and quantifies the value of a customer to the organization.
* Behavior-based segmentation: Analyzes customer product usage and purchasing patterns.

**Answer 8:** Regression predicts a continuous dependent element in the presence of various independent elements. Linear regression tries to make a trend line that has the least difference between actual and predicted values. This difference is also known as residual.

RESIDUAL = ACTUAL VALUE - PREDICTED VALUE

Before evaluating the model using evaluation metrics, one should go for a residual plot.

Residuals are significant when figuring the quality of the model. One should look at two things in residuals, their magnitude and whether they form a pattern or not.

* The further away the residuals are from 0, the more faulty the model is.
* When the residual’s average is not zero, it shows that the model is frequently under or over predicting.
* When the residual plot consists of patterns, it shows the model fails to explain a few properties of the data.

WHAT DOES A RESIDUAL PLOT DO?

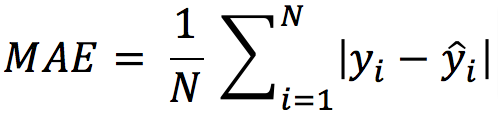
Residual plots tell us if the model is biased or not, better than any other performance metric. If the residual plot shows a random distribution of the predicted and actual values and not a particular pattern, then we can evaluate the model with various other metrics.

We have residuals for each and every data point, but we need some summarizing agents to see the overview of the data. That’s where these metrics come in.

The metrics to evaluate regression models are as follows:

MEAN ABSOLUTE ERROR (MAE):

It is the mean of the absolute difference between the actual value in the dataset and the value predicted by the model.

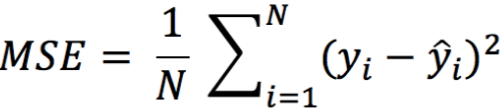


N = the count of the data points.  
Y = the actual value in the dataset  
Y cap = the model’s predicted value.

The absolute values are taken, and if it’s not then the negative and positive difference will cancel out each other. The smaller the MAE, the more accurate the model is. If MAE is zero it shows the model is perfect. If MAE is large then the model is not good.

MEAN SQUARED ERROR (MSE):

This is the mean of the squared difference of the actual value in the dataset and the value predicted by the model.



N = the count of data points in the data.

Y = the actual value in the dataset.

Y cap=  the model’s predicted value.

The MSE will be large if there are outliers in the dataset, this is not the case with MAE.

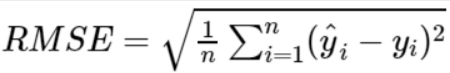
* MSE focuses on larger errors, as when we are squaring the error the effect of large errors becomes more prominent.
* If the errors are low, lower than one, then it leads to underestimating the model’s error.

ROOT MEAN SQUARED ERROR:

It is the mean of root squared subtraction between the actual value in the dataset and the value predicted by the model.

It’s the same as MSE, we are just taking the root of it.

The smaller the value of root mean squared error, the more accurate the model is.

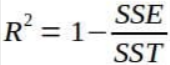


\*\*R Square: \*\*

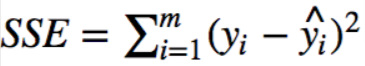
It estimates the ratio of the variance of the dependent element described by the target element.

It’s used for finding the accuracy of the model. It depicts the closeness of the data points to the trend line made by the model. This helps to make a link between the independent element and the target element.

R square is from zero to one. The nearer R square is to one, the more accurate the model.



SSE is the sum of the square difference of the residuals.



SST is the sum of the difference of the actual value of the data and the mean of all the actual values in data.

image

When we add more elements in the data then R square increases, for this we can use adjusted R square.

I hope you now have ideas of different metrics of evaluating a regression model.