If your lambda value is too high, your model will be simple, but you run the risk of underfitting your data. Your model won't learn enough about the training data to make useful predictions.

AdaBoost algorithm, is a Boosting technique used as an Ensemble Method in Machine Learning. The most common algorithm used with AdaBoost is decision trees with one level that means with Decision trees with only 1 split. it builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. Now all the points which have higher weights are given more importance in the next model. It will keep training models until and unless a lowe error is received.

Bayesian classification is based on Bayes' Theorem. Bayesian classifiers are the statistical classifiers. Bayesian classifiers can predict class membership probabilities such as the probability that a given tuple belongs to a particular class.

Bias is a phenomenon that skews the result of an algorithm in favor or against an idea. Bias is considered a systematic error that occurs in the machine learning model itself due to incorrect assumptions in the ML process. Bias in ML does help us generalize better and make our model less sensitive to some single data point. Low Bias: A low bias model will make fewer assumptions about the form of the target function. High Bias: A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. A high bias model also cannot perform well on new data.

Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data.

* Logistic Regression
* Naive Bayes classification algorithm is a probabilistic classifier. It is based on probability models that incorporate strong independence assumptions.
* SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
* Decision Trees, Random Forest, Neural network

Confusion Matrix is a two by two table that contains four outcomes produced by a binary classifier. Various measures, such as error-rate, accuracy, specificity, sensitivity, and precision, are derived from the confusion matrix.

Elastic net is a penalized linear regression model that includes both the L1 and L2 penalties during training.

The Gini index (or coefficient) is a synthetic indicator that captures the level of inequality for a given variable and population. It varies between 0 (perfect equality) and 1 (extreme inequality).

When the data is linearly separable, and we don’t want to have any misclassifications, we use SVM with a hard margin. The constraint of maximizing the margin of the line that separates the classes must be relaxed. This is often called the soft margin

Gradient descent (GD) is an iterative first-order optimisation algorithm used to find a local minimum/maximum of a given function. This method is commonly used in machine learning (ML) and deep learning(DL) to minimise a cost/loss function

Batch gradient descent sums the error for each point in a training set, updating the model only after all training examples have been evaluated. advantages of batch gradient descent are its computational efficient, it produces a stable error gradient and a stable convergence. Some disadvantages are the stable error gradient can sometimes result in a state of convergence that isn’t the best the model can achieve.

Kernel trick allows the inner product of mapping function instead of the data points. The trick is to identify the kernel functions which can be represented in place of the inner product of mapping functions. Kernel functions allow easy computation.

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process

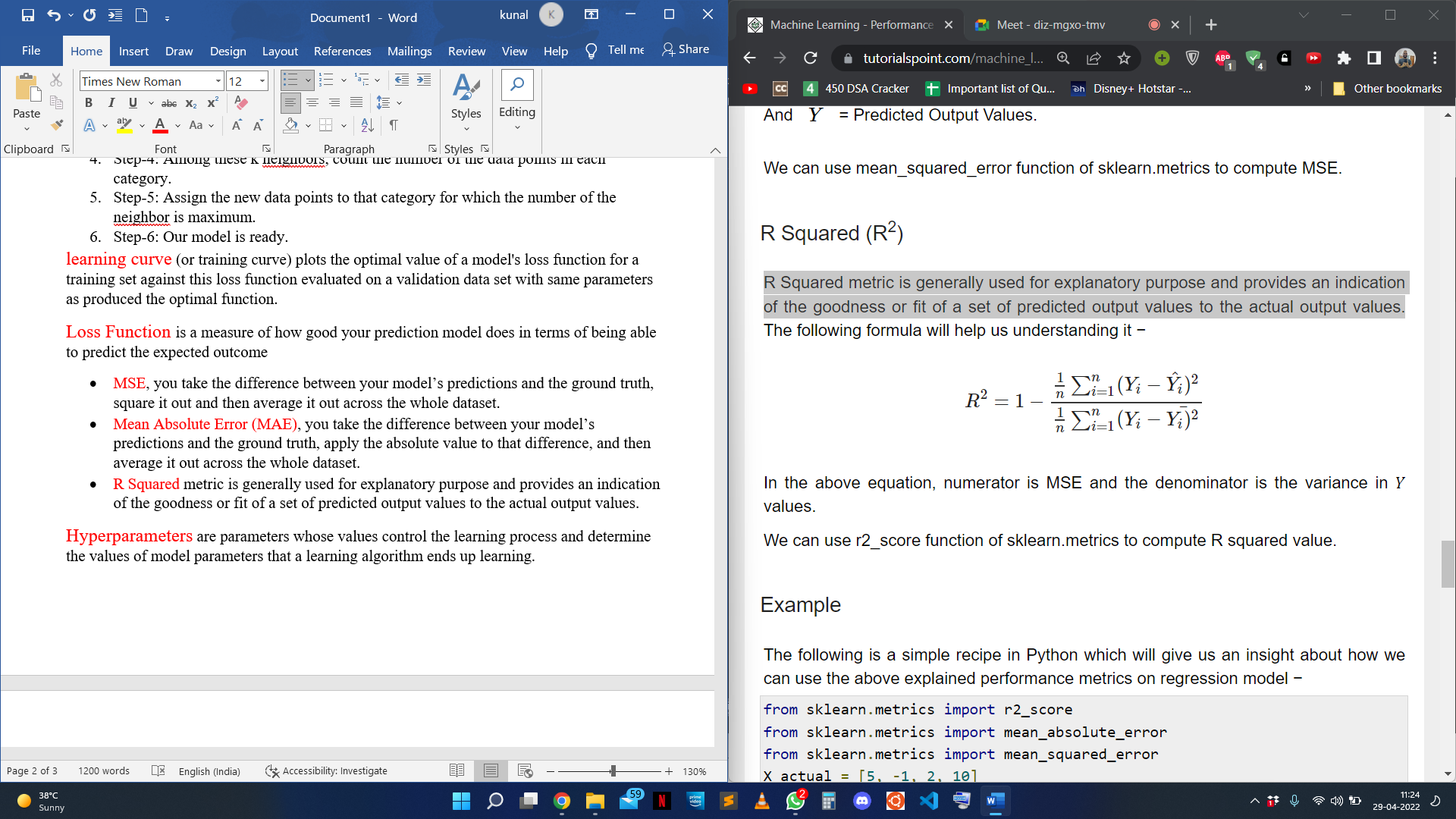
1. Partition objects into k non-empty subsets
2. Compute seed points as the centroids of the clusters of the current partition (the centroid is the center, i.e., mean point, of the cluster)
3. Assign each object to the cluster with the nearest seed point
4. Go back to Step 2, stop when no more new assignment

K-NN algorithm is the simplest Machine Learning algorithms based on Supervised Learning technique. can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It classifies the data point on how its neighbour is classified. KNN classifies the new data points based on the similarity measure of the earlier stored data points.

1. Step-1: Select the number K of the neighbors
2. Step-2: Calculate the Euclidean distance of K number of neighbors
3. Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
4. Step-4: Among these k neighbors, count the number of the data points in each category.
5. Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
6. Step-6: Our model is ready.

learning curve (or training curve) plots the optimal value of a model's loss function for a training set against this loss function evaluated on a validation data set with same parameters as produced the optimal function.

Loss Function is a measure of how good your prediction model does in terms of being able to predict the expected outcome

* MSE, you take the difference between your model’s predictions and the ground truth, square it out and then average it out across the whole dataset.
* Mean Absolute Error (MAE), you take the difference between your model’s predictions and the ground truth, apply the absolute value to that difference, and then average it out across the whole dataset.
* R Squared metric is generally used for explanatory purpose and provides an indication of the goodness or fit of a set of predicted output values to the actual output values.

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.

Regression is a method to determine the statistical relationship between a dependent variable and one or more independent variables. It has continuous output.

* Linear regression
* Decision Trees
* Random Learning
* Neural Network

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. The commonly used regularization techniques are :

* L1 regularization
* L2 regularization
* Dropout regularization

A regression model that uses L1 regularization technique is called Lasso Regression and model which uses L2 is called Ridge Regression. Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function. Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds “absolute value of magnitude” of coefficient as penalty term to the loss function. The key difference between these techniques is that Lasso shrinks the less important feature’s coefficient to zero thus, removing some feature altogether. So, this works well for feature selection in case we have a huge number of features.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane.

Stacking is an ensemble machine learning algorithm that learns how to best combine the predictions from multiple well-performing machine learning models. Stacking is a general procedure where a learner is trained to combine the individual learners. Here, the individual learners are called the first-level learners, while the combiner is called the second-level learner, or meta-learner.