*Question no 1*

*1.Eigen (pronounced eye-gan) is a German word that means “own” or “innate”, as in belonging to the parent matrix.*

*2.Eigenvectors are unit vectors, which means that their length or magnitude is equal to 1. They are often referred to as right vectors, whichsimply means a column vector.Whereas, eigenvalues are coefficients applied to eigenvectors that give the vectors their length or magnitude.*

*3.Eigendecomposition is used to decompose a matrix into eigenvectors and eigenvalues which are eventually applied in methods used in machine learning, such as in the Principal Component Analysis method or PCA.*

*4. Decomposing a matrix in terms of its eigenvalues and its eigenvectors gives valuable insights into the properties of the matrix.*

*5.Machine learning involves lots of data. If in any way we are able to reduce the data size without losing actual representation of those values then its a piece of pie for programmers.*

*6.Principal Component Analysis or PCA is performed for dimensionality reduction. With larger datasets, finding significant features gets difficult.*

*So, in order to check for the correlation between two variables and if they could be dropped off the table to make the machine learning model more robust.*

*Question no 2*

*Data imbalance usually reflects an unequal distribution of classes within a dataset.*

*For example, in a credit card fraud detection dataset, most of the credit card transactions are not fraud and a very few classes are fraud transactions.*

*This leaves us with something like 50:1 ratio between the fraud and non-fraud classes.*

*\*\* BY using some techniques we can resolve this issue*

*1- Resampling (Oversampling and Undersampling):*

*Undersampling is the process where you randomly delete some of the observations from the majority class in order to match the numbers with the minority class.*

*The second resampling technique is called, Oversampling. This process is a little more complicated than undersampling.*

*It is the process of generating synthetic data that tries to randomly generate a sample of the attributes from observations in the minority class.*

*There are a number of methods used to oversample a dataset for a typical classification problem. The most common technique is called SMOTE (Synthetic Minority Over-sampling Technique).*

*In simple terms, it looks at the feature space for the minority class data points and considers its k nearest neighbours.*

*2- Ensembling Methods (Ensemble of Sampler):*

*In Machine Learning, ensemble methods use multiple learning algorithms and techniques to obtain better performance than what could be obtained from any of the constituent learning algorithms alone.*

*(Yes, just like a democratic voting system). When using ensemble classifiers, bagging methods become popular and it works by building multiple estimators on a different randomly selected subset of data.*

*In the scikit-learn library, there is an ensemble classifier namedBaggingClassifier. However, this classifier does not allow to balance each subset of data.*

*Therefore, when training on imbalanced data set, this classifier will favour the majority classes and create a biased model.*

*Question no 3*

*First we will see what is a bias and variance*

*\*\*\*Bias :- Bias is the difference between the average prediction of our model and the correct value which we are trying to predict.*

*Model with high bias pays very little attention to the training data and oversimplifies the model.*

*It always leads to high error on training and test data.*

*\*\*\*variance :- Variance is the variability of model prediction for a given data point or a value which tells us spread of our data.*

*Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn’t seen before.*

*As a result, such models perform very well on training data but has high error rates on test data.*

*\*\*\* Bias-Variance trade-off*

*If the algorithm is too simple (hypothesis with linear eq.) then it may be on high bias and low variance condition and thus is error-prone.*

*If algorithms fit too complex ( hypothesis with high degree eq.) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well.*

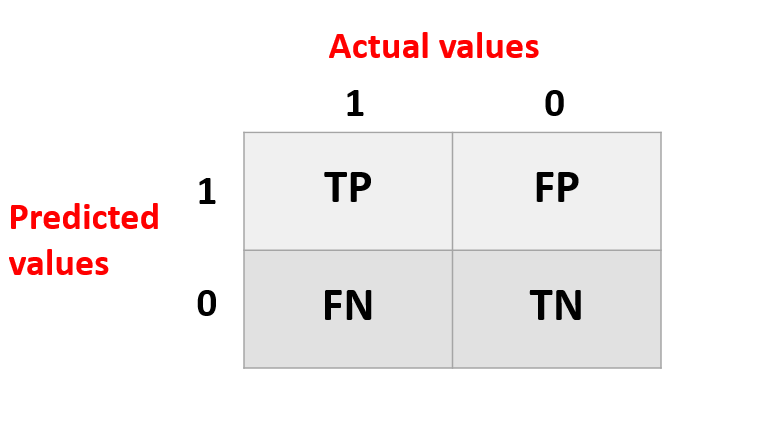
*Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off.*

*Question no 4*

**What is Confusion Matrix?**

Confusion Matrix is the visual representation of the Actual VS Predicted values. It measures the performance of our Machine Learning classification model and looks like a table-like structure.

This is how a Confusion Matrix of a binary classification problem looks like :



**Elements of Confusion Matrix**

It represents the different combinations of Actual VS Predicted values. Let’s define them one by one.

**TP: True Positive:** The values which were actually positive and were predicted positive.

**FP: False Positive:** The values which were actually negative but falsely predicted as positive. Also known as Type I Error.

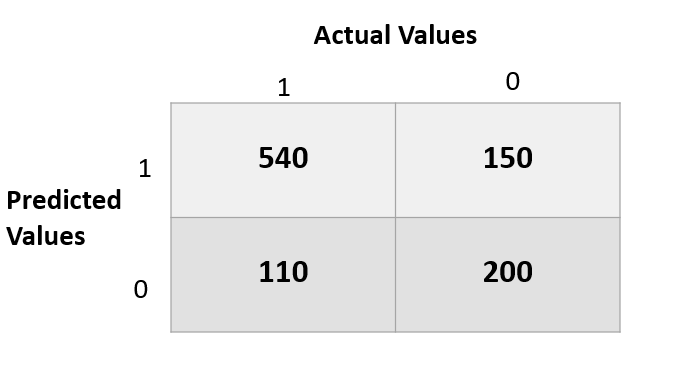
**FN: False Negative**: The values which were actually positive but falsely predicted as negative. Also known as Type II Error.

**TN: True Negative:** The values which were actually negative and were predicted negative.

**Understanding it with the help of an example**

Taking an example of the **Stock Market Crash prediction** project. This is a binary classification problem where **1 means the stock market will crash** and **0 means the stock market will not crash and**suppose we have **1000 records** in our dataset.

Let’s see the confusion matrix of the following :



In the above matrix, we can analyze the model as :

True positive: 540 records of the stock market crash were **predicted correctly** by the model.

False-positive: 150 records of not a stock market crash were **wrongly predicted** as a market crash.

False-negative: 110 records of a market crash were **wrongly predicted**as not a market crash.

True Negative: 200 records of not a market crash were predicted correctly by the model.

**Other Evaluation Metrics associated with it**

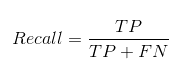
Accuracy:

It is calculated by dividing the total number of correct predictions by all the predictions.

Equation_Accuracy

Recall / Sensitivity:

The recall is the measure to check correctly positive predicted outcomes out of the total number of positive outcomes.



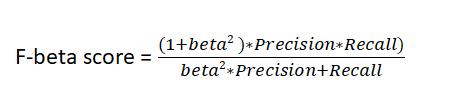
Precision:

Precision checks how many outcomes are actually positive outcomes out of the total positively predicted outcomes.

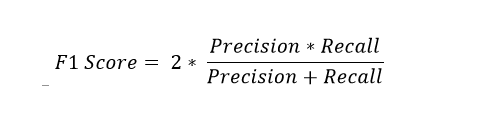
Confusion matrix Precision

F beta score:

F beta score is the harmonic mean of Precision and Recall and it captures the contribution of both of them. The contribution depends on the beta value in the below formula.



The default beta value is 1 which gives us the formula of F1score, where the contribution of Precision and Recall are the same. Higher the F1 score, the better the model.



The beta value < 1 gives more weight to Precision than Recall and the beta value>1 gives more weight to Recall.

Question no 5

Simple linear regression is a regression technique in which the independent variable has a linear relationship with the dependent variable. The straight line in the diagram is the best fit line. The main goal of the simple linear regression is to consider the given data points and plot the best fit line to fit the model in the best way possible.

\*\*Drawbacks of linear models

1. The assumption of linearity between dependent and independent variables
2. It is often quite prone to noise and overfitting
3. Linear regression is quite sensitive to outliers
4. It is prone to multicollinearity

Question no 6

**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 (e.g. in a linear regression). Due to its importance and ease of implementation, this algorithm is usually taught at the beginning of almost all machine learning courses.

\*\*Gradient :- In the case of**a univariate function**, it is simply the **first derivative at a selected point**. In the case of **a multivariate function**, it is a **vector of derivatives** in each main direction (along variable axes). Because we are interested only in a slope along one axis and we don’t care about others these derivatives are called **partial derivatives.**

# Gradient Descent Algorithm

Gradient Descent Algorithm iteratively calculates the next point using gradient at the current position, scales it (by a learning rate) and subtracts obtained value from the current position (makes a step). It subtracts the value because we want to minimise the function (to maximise it would be adding). This process can be written as:



There’s an important parameter **η** which scales the gradient and thus controls the step size. In machine learning, it is called **learning rate** and have a strong influence on performance.

* The smaller learning rate the longer GD converges, or may reach maximum iteration before reaching the optimum point
* If learning rate is too big the algorithm may not converge to the optimal point (jump around) or even to diverge completely.

In summary, Gradient Descent method’s steps are:

1. choose a starting point (initialisation)
2. calculate gradient at this point
3. make a scaled step in the opposite direction to the gradient (objective: minimise)
4. repeat points 2 and 3 until one of the criteria is met

Question no 7

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the 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. This best decision boundary is called a hyperplane.