Types of Regression used in Machine Learning.

Linear Regression :

Linear Regression is the most basic approach for prediction analysis of data as it is applicable to all domains and fields. The Linear Regression Model consists of the predictor value and a dependent variable related linearly to each other. In case the data involves more than one independent variable, then linear regression is called multivariate-linear-regression.

The following equation is used to represent the Linear Regression model :

$y = mx+c+e$

where m is the slope, c is the intercept, and e is the standard error in the model.

When : When we want to predict the value of one variable using another variable.

Why : Analysts use Linear Regression to evaluate trends in business, weather forecasts, estimating market value and many more real-world applications.

The best fit line is determined by varying the values of m and c. The predictor error is the difference between the observed values and the predicted value. The values of m and c get selected in such a way that it gives minimum possible error. It is also important to note that a linear regression model is susceptible to outliers. Therefore, it should not be used in the case of big size data.

Steps to follow Linear Regression:

- Import the libraries and dataset.
- Check the dataset for Nan values, evaluate the data points.
- Clean the dataset before further processing.
- Decide your independent and dependent variables and calculate the sizes.
- Find out the means of both x and y.
- Calculate cross deviation and deviation, sum of squared errors and the regression coefficients.
- Plot the points (Scatter)
- Predict the regression line value, using the minimum error possible.
- Plot the regression line and find the labels.
- Display complete graph.

Advantages	**Drawbacks**
Simplicity	Inaccuracy in many real-world phenomena.
Interpretability	Assumptions may not correspond to the actual values.
Scientific Acceptance	
Widespread	

Logistic Regression :

Logistic Regression is a machine learning algorithm which is used for the classification problems, it is a predictive analysis algorithm based on the concept of probability.

We can call Logistic Regression a Linear Regression model, but the Logistic Regression uses a more complex cost function can be defined as the "Sigmoid Function" or also known as Logistic Function instead of a linear function.

The hypothesis of logistic regression, unlike linear regression tends it to limit the cost function between 0 and 1. Therefore linear functions fail to represent it as it can have a value greater than 1 or less than 0 which is possible as per the hypothesis of logistic regression.

When : Logistic regression is used where there is a need for classification of data.

Why : Logistic regression keeps the prediction value between 0 and 1 so we use it when we must find the stage of the process (out of 100%) or because we are in need of classification. (True or False based).

y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x))

Where y is the predicted output, b0 is the bias or intercept term and b1 is the coefficient for the single input value (x). Each column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data.

Sigmoid function – In order to map predicted values to probabilities, we use the Sigmoid function. The function maps any real value into another value between 0 and 1. In machine learning, we use sigmoid to map predictions to probabilities.

Steps to performing Logistic Regression :

- Importing the libraries and dataset.
- Pre-processing the dataset.
- Fitting Logistic Regression to the Training set.
- Predicting the test result.
- Test Accuracy of the result.
- Plot the confusion matrix.
- Visualizing the test set result.

Advantages	**Drawbacks**
Easy to implement, interpret and train.	Number of observations is less than features, may lead to overfitting.
Can easily be used in multinomial and a natural probabilistic view of class predictions.	Requires next to no multicollinearity between independent variables.
Can interpret model coefficients as indicators of feature importance.	Non-linear problems can't be solved with logistic regression because it has a linear decision surface
It not only provides a measure of how appropriate a predictor(coefficient size)is, but also its direction of association (positive or negative)	It can only be used to predict discrete functions
Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.	It constructs linear boundaries.

Ridge Regression :

The Ridge regression is used when there is a high corelation between the independent variables. This is because, in the case of multi collinear data, the least square estimates give unbiased values. But, in case the collinearity is high, there can be a bias value.

Therefore, a bias matrix is introduced in the equation of ridge regression. This is a very powerful regression method where the model is susceptible to overfitting.

The equation used to denote RR :

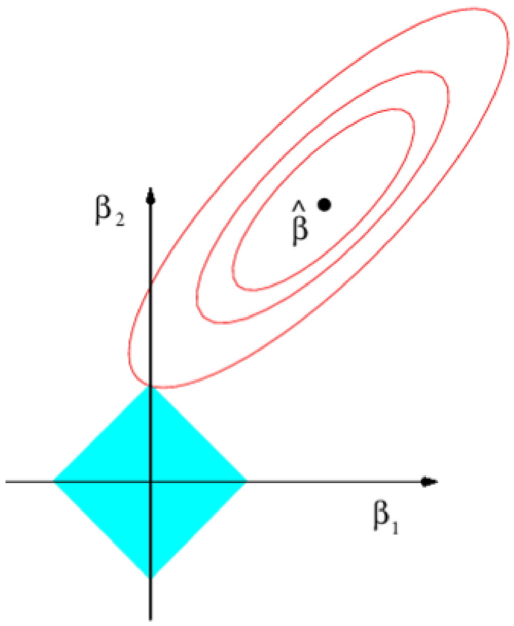
$$\beta = (X^{T}X + \lambda * I)^{-1}X^{T}y$$

When : Ridge regression is used where there is multicollinearity, least squares are unbiased ,variance is large.

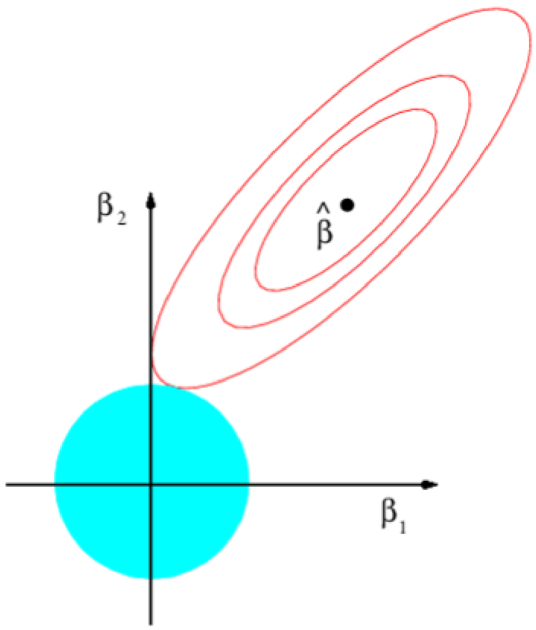
Why : It is used to analyse the data that suffers from multicollinearity.

Steps to perform Ridge Regression :

- Importing libraries and dataset.
- Pre-Processing the dataset.
- Train-Test split.
- Creating a Linear Regression model and finding the best fit model on the training data.
- Perform regularization, Value of alpha, which is a hyperparameter of Ridge, which means that they are not automatically learned by the model instead they must be set manually. We run a grid search for optimum alpha values, to find optimum alpha for Ridge Regularization we are applying GridSearchCV.
- Plotting the data.



Lasso



Ridge

Lasso Regression :

Lasso Regression is the type of regression that performs regularization along with feature selection. it prohibits the absolute size of the regression coefficient. As a result, the coefficient value gets nearer to zero, which does not happen in the case of ridge regression because we are nullifying the problem of multicollinearity.

Due to this, feature selection is used in Lasso, which allows selecting a set of features from the dataset to build the model. In the case of Lasso, only the required features are used, and the other ones are made zero, to maintain the shape of the graph and negate any spike factors.

The goal of the algorithm is to minimize :

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Which is somewhat on the same lines as minimizing the sum of squares just with constraint Lim $B_j < s$.

Some of the B_s are shrunk to zero.

A tuning parameter λ controls the strength of the L1 penalty. λ is basically the amount of shrinkage.

- When λ = 0, no parameters are eliminated. The estimate is equal to the one found with linear regression.
- As λ increases, more and more coefficients are set to zero and eliminated (theoretically, when λ = ∞, all coefficients are eliminated).
- As λ increases, bias increases.
- As λ decreases, variance increases.

Advantages and Disadvantages of using regularization methods like Ridge and Lasso –

Advantages

- Avoids overfitting a model.
- They do not require unbiased estimators.
- They add just enough bias to make the estimates reasonably reliable approximations to true population values.
- They still perform well in cases of a large multivariate data with the number of predictors (p) larger than the number of observations (n).
- The ridge estimator is preferably good at improving the least-squares estimate when there is multicollinearity.

Disadvantages

- They include all the predictors in the final model.
- They are unable to perform feature selection.
- They shrink the coefficients towards zero.
- They trade the variance for bias.

Bayesian Linear Regression :

Bayesian linear regression allows a natural mechanism to survive insufficient data, or poor distributed data. It allows you to put a prior on the coefficients and on the noise so that in the absence of data, the priors can take over. More importantly, you can ask Bayesian linear regression which parts (if any) of its fit to the data is it confident about, and which parts are very uncertain (perhaps based entirely on the priors).

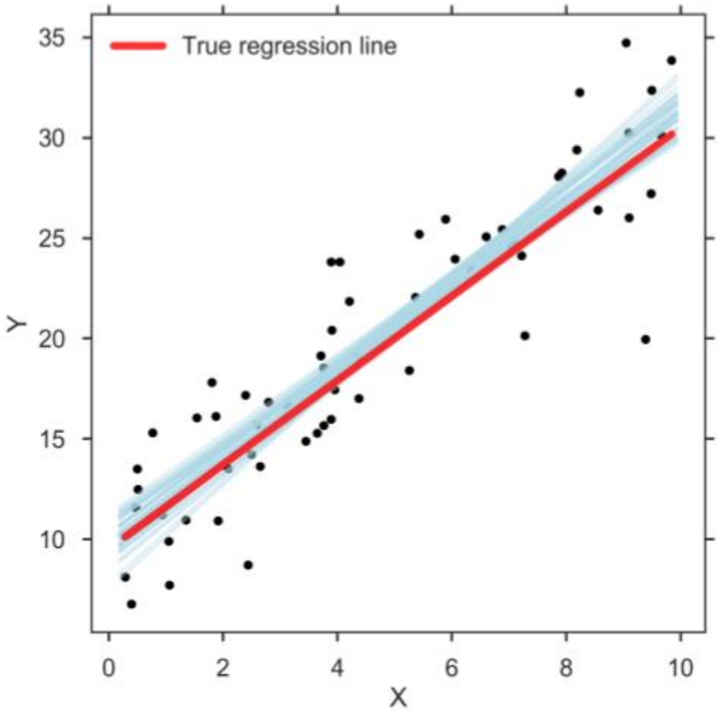
$$P(A \mid B) = \frac{P(B|A)\, P(A)}{P(B)}$$

Advantages of Bayesian Regression:

- Very effective when the size of the dataset is small.
- Particularly well-suited for on-line based learning (data is received in real-time), as compared to batch-based learning, where we have the entire dataset on our hands before we start training the model. This is because Bayesian Regression does not need to store data.
- The Bayesian approach is a tried and tested approach and is very robust, mathematically. So, one can use this without having any extra prior knowledge about the dataset.

Disadvantages of Bayesian Regression:

- The inference of the model can be time-consuming.
- If there is a large amount of data available for our dataset, the Bayesian approach is not worth it and the regular frequentist approach does a more efficient job.



Programmatical Approach for each method of Regression:

Linear Regression :

https://github.com/joshtrivedi/Machine-Learning/blob/main/Assignments/Types-Of-Regression/Basic_Linear_Regression.ipynb

Logistic Regression :

https://github.com/joshtrivedi/Machine-Learning/blob/main/Assignments/Types-Of-Regression/Logistic_regression.ipynb

Ridge Regression :

https://github.com/joshtrivedi/Machine-Learning/blob/main/Assignments/Types-Of-Regression/Ridge_Regression.ipynb

Lasso Regression :

https://github.com/joshtrivedi/Machine-Learning/blob/main/Assignments/Types-Of-Regression/Lasso_Regression.ipynb

Bayesian Linear Regression :

https://github.com/joshtrivedi/Machine-Learning/blob/main/Assignments/Types-Of-Regression/Bayesian.py