Linear Regression in Machine learning

Machine Learning is a branch of Artificial intelligence that focuses on the development of algorithms and statistical models that can learn from and make predictions on data. **Linear regression** is also a type of machine-learning algorithm more specifically a **supervised machine-learning algorithm** that learns from the labelled datasets and maps the data points to the most optimized linear functions, which can be used for prediction on new datasets.

First off we should know what supervised machine learning algorithms is. It is a type of machine learning where the algorithm learns from labelled data. Labeled data means the dataset whose respective target value is already known. Supervised learning has two types:

- **Classification**: It predicts the class of the dataset based on the independent input variable. Class is the categorical or discrete values. like the image of an animal is a cat or dog?
- **Regression**: It predicts the continuous output variables based on the independent input variable. like the prediction of house prices based on different parameters like house age, distance from the main road, location, area, etc.

What is Linear Regression?

Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between the dependent variable and one or more independent features by fitting a linear equation to observed data.

When there is only one independent feature, it is known as Simple Linear Regression, and when there are more than one feature, it is known as Multiple Linear Regression.

Similarly, when there is only one dependent variable, it is considered Univariate Linear Regression, while when there are more than one dependent variables, it is known as Multivariate Regression.

Why Linear Regression is Important?

The interpretability of linear regression is a notable strength. The model's equation provides clear coefficients that elucidate the impact of each independent variable on the dependent variable, facilitating a deeper understanding of the underlying dynamics. Its simplicity is a virtue, as linear regression is transparent, easy to implement, and serves as a foundational concept for more complex algorithms.

Linear regression is not merely a predictive tool; it forms the basis for various advanced models. Techniques like regularization and support vector machines draw inspiration from linear regression, expanding its utility. Additionally, linear regression is a cornerstone in assumption testing, enabling researchers to validate key assumptions about the data.

Types of Linear Regression

There are two main types of linear regression:

Simple Linear Regression

This is the simplest form of linear regression, and it involves only one independent variable and one dependent variable. The equation for simple linear regression is:
$y=\beta_0+\beta_1X y=\beta_0+\beta_1X$
where:

- Y is the dependent variable
- X is the independent variable
- β_0 is the intercept
- β_1 is the slope

Multiple Linear Regression

This involves more than one independent variable and one dependent variable. The equation for multiple linear regression is:
$y=\beta_0+\beta_1X_1+\beta_2X_2+\dots\dots\dots\beta_nX_n y=\beta_0+\beta_1X_1+\beta_2X_2+\dots\dots\dots\beta_nX_n$
where:

- Y is the dependent variable
- $X_1, X_2, \dots, X_n$ are the independent variables
- β_0 is the intercept
- $\beta_1, \beta_2, \dots, \beta_n$ are the slopes

- The goal of the algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.
- In regression set of records are present with X and Y values and these values are used to learn a function so if you want to predict Y from an unknown X this learned function can be used. In regression we have to find the value of Y, So, a function is required that predicts continuous Y in the case of regression given X as independent features.

What is the best Fit Line?

Our primary objective while using linear regression is to locate the best-fit line, which implies that the error between the predicted and actual values should be kept to a minimum. There will be the least error in the best-fit line.

The best Fit Line equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable(s).

Here Y is called a dependent or target variable and X is called an independent variable also known as the predictor of Y. There are many types of functions or modules that can be used for regression. A linear function is the simplest type of function. Here, X may be a single feature or multiple features representing the problem.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x)). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best-fit line for our model.

We utilize the cost function to compute the best values in order to get the best fit line since different values for weights or the coefficient of lines result in different regression lines.

Hypothesis function in Linear Regression

As we have assumed earlier that our independent feature is the experience i.e X and the respective salary Y is the dependent variable. Let's assume there is a linear relationship between X and Y then the salary can be predicted using:

$$\hat{Y} = \theta_1 + \theta_2 X$$

OR

$$\hat{y}_i = \theta_1 + \theta_2 x_i$$

The model gets the best regression fit line by finding the best θ_1 and θ_2 values.

- **θ_1:** intercept
- **θ_2:** coefficient of x

Once we find the best θ_1 and θ_2 values, we get the best-fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update θ_1 and θ_2 values to get the best-fit line?

To achieve the best-fit regression line, the model aims to predict the target value $\hat{Y}$ such that the error difference between the predicted value $\hat{Y}$ and the true value Y is minimum. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimizes the error between the predicted y value (pred) and the true y value (y).

$$minimize \frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2$$

Cost function for Linear Regression

The cost function or the loss function is nothing but the error or difference between the predicted value Y^ *Y^* and the true value Y.

In Linear Regression, the **Mean Squared Error (MSE)** cost function is employed, which calculates the average of the squared errors between the predicted values y^i*y^i* and the actual values yi*yi*. The purpose is to determine the optimal values for the intercept θ1*θ*1 and the coefficient of the input feature θ2*θ*2 providing the best-fit line for the given data points. The linear equation expressing this relationship is y^i=θ1+θ2xi*y^i* =*θ*1+*θ*2*xi*.

MSE function can be calculated as:

Cost function(J)=1n∑ni(yi^−yi)2Cost function(*J*)=*n*1∑*ni*(*yi*^−*yi*)2

Utilizing the MSE function, the iterative process of gradient descent is applied to update the values of \θ1&θ2*θ*1&*θ*2. This ensures that the MSE value converges to the global minima, signifying the most accurate fit of the linear regression line to the dataset.

This process involves continuously adjusting the parameters \(\theta_1\) and \(\theta_2\) based on the gradients calculated from the MSE. The final result is a linear regression line that minimizes the overall squared differences between the predicted and actual values, providing an optimal representation of the underlying relationship in the data.

Gradient Descent for Linear Regression

A linear regression model can be trained using the optimization algorithm gradient descent by iteratively modifying the model's parameters to reduce the mean squared error (MSE) of the model on a training dataset. To update θ1 and θ2 values in order to reduce the Cost function (minimizing RMSE value) and achieve the best-fit line the model uses Gradient Descent. The idea is to start with random θ1 and θ2 values and then iteratively update the values, reaching minimum cost.

Assumptions of Simple Linear Regression

Linear regression is a powerful tool for understanding and predicting the behavior of a variable, however, it needs to meet a few conditions in order to be accurate and dependable solutions.

1. **Linearity**: The independent and dependent variables have a linear relationship with one another. This implies that changes in the dependent variable follow those in the independent variable(s) in a linear fashion. This means that there should be a straight line that can be drawn through the data points. If the relationship is not linear, then linear regression will not be an accurate model.
2. **Independence**: The observations in the dataset are independent of each other. This means that the value of the dependent variable for one observation does not depend on the value of the dependent variable for another observation. If the

observations are not independent, then linear regression will not be an accurate model.

3. **Homoscedasticity**: Across all levels of the independent variable(s), the variance of the errors is constant. This indicates that the amount of the independent variable(s) has no impact on the variance of the errors. If the variance of the residuals is not constant, then linear regression will not be an accurate model.
4. Normality: The residuals should be normally distributed. This means that the residuals should follow a bell-shaped curve. If the residuals are not normally distributed, then linear regression will not be an accurate model.

Assumptions of Multiple Linear Regression

For Multiple Linear Regression, all four of the assumptions from Simple Linear Regression apply. In addition to this, below are few more:

1. **No multicollinearity**: There is no high correlation between the independent variables. This indicates that there is little or no correlation between the independent variables. Multicollinearity occurs when two or more independent variables are highly correlated with each other, which can make it difficult to determine the individual effect of each variable on the dependent variable. If there is multicollinearity, then multiple linear regression will not be an accurate model.
2. **Additivity:** The model assumes that the effect of changes in a predictor variable on the response variable is consistent regardless of the values of the other variables. This assumption implies that there is no interaction between variables in their effects on the dependent variable.
3. **Feature Selection:** In multiple linear regression, it is essential to carefully select the independent variables that will be included in the model. Including irrelevant or redundant variables may lead to overfitting and complicate the interpretation of the model.
4. **Overfitting:** Overfitting occurs when the model fits the training data too closely, capturing noise or random fluctuations that do not represent the true underlying relationship between variables. This can lead to poor generalization performance on new, unseen data.

Multicollinearity

Multicollinearity is a statistical phenomenon that occurs when two or more independent variables in a multiple regression model are highly correlated, making it difficult to assess the individual effects of each variable on the dependent variable.

Detecting Multicollinearity includes two techniques:

- **Correlation Matrix:** Examining the correlation matrix among the independent variables is a common way to detect multicollinearity. High correlations (close to 1 or -1) indicate potential multicollinearity.

- **VIF (Variance Inflation Factor):** VIF is a measure that quantifies how much the variance of an estimated regression coefficient increases if your predictors are correlated. A high VIF (typically above 10) suggests multicollinearity.

Evaluation Metrics for Linear Regression

A variety of evaluation measures can be used to determine the strength of any linear regression model. These assessment metrics often give an indication of how well the model is producing the observed outputs.

The most common measurements are:

Mean Square Error (MSE)

Mean Squared Error (MSE) is an evaluation metric that calculates the average of the squared differences between the actual and predicted values for all the data points. The difference is squared to ensure that negative and positive differences don't cancel each other out.

MSE is a way to quantify the accuracy of a model's predictions. MSE is sensitive to outliers as large errors contribute significantly to the overall score.

Mean Absolute Error (MAE)

Mean Absolute Error is an evaluation metric used to calculate the accuracy of a regression model. MAE measures the average absolute difference between the predicted values and actual values.

Mathematically, MAE is expressed as:

Lower MAE value indicates better model performance. It is not sensitive to the outliers as we consider absolute differences.

Root Mean Squared Error (RMSE)

The square root of the residuals' variance is the Root Mean Squared Error. It describes how well the observed data points match the expected values, or the model's absolute fit to the data.

RSME is not as good of a metric as R-squared. Root Mean Squared Error can fluctuate when the units of the variables vary since its value is dependent on the variables' units (it is not a normalized measure).

Coefficient of Determination **(R-squared)**

R-Squared is a statistic that indicates how much variation the developed model can explain or capture. It is always in the range of 0 to 1. In general, the better the model matches the data, the greater the R-squared number.

R squared metric is a measure of the proportion of variance in the dependent variable that is explained the independent variables in the model.

Adjusted R-Squared Error

Adjusted R2 measures the proportion of variance in the dependent variable that is explained by independent variables in a regression model. Adjusted R-square accounts the number of predictors in the model and penalizes the model for including irrelevant predictors that don't contribute significantly to explain the variance in the dependent variables.

Adjusted R-square helps to prevent overfitting. It penalizes the model with additional predictors that do not contribute significantly to explain the variance in the dependent variable.

Linear Regression Line

The linear regression line provides valuable insights into the relationship between the two variables. It represents the best-fitting line that captures the overall trend of how a dependent variable (Y) changes in response to variations in an independent variable (X).

- **Positive Linear Regression Line**: A positive linear regression line indicates a direct relationship between the independent variable (X) and the dependent variable (Y). This means that as the value of X increases, the value of Y also increases. The slope of a positive linear regression line is positive, meaning that the line slants upward from left to right.
- **Negative Linear Regression Line**: A negative linear regression line indicates an inverse relationship between the independent variable (X) and the dependent variable (Y). This means that as the value of X increases, the value of Y decreases. The slope of a negative linear regression line is negative, meaning that the line slants downward from left to right.

Regularization Techniques for Linear Models

Lasso Regression (L1 Regularization)

Lasso Regression is a technique used for regularizing a linear regression model, it adds a penalty term to the linear regression objective function to prevent overfitting.

Ridge Regression (L2 Regularization)

Ridge regression is a linear regression technique that adds a regularization term to the standard linear objective. Again, the goal is to prevent overfitting by penalizing large coefficient in linear regression equation. It useful when the dataset has multicollinearity where predictor variables are highly correlated.

Elastic Net Regression

Elastic Net Regression is a hybrid regularization technique that combines the power of both L1 and L2 regularization in linear regression objective

Applications of Linear Regression

Linear regression is used in many different fields, including finance, economics, and psychology, to understand and predict the behavior of a particular variable. For example, in finance, linear regression might be used to understand the relationship between a company's stock price and its earnings or to predict the future value of a currency based on its past performance.

Advantages & Disadvantages of Linear Regression

Advantages of Linear Regression

- Linear regression is a relatively simple algorithm, making it easy to understand and implement. The coefficients of the linear regression model can be interpreted as the change in the dependent variable for a one-unit change in the independent variable, providing insights into the relationships between variables.
- Linear regression is computationally efficient and can handle large datasets effectively. It can be trained quickly on large datasets, making it suitable for real-time applications.
- Linear regression is relatively robust to outliers compared to other machine learning algorithms. Outliers may have a smaller impact on the overall model performance.
- Linear regression often serves as a good baseline model for comparison with more complex machine learning algorithms.
- Linear regression is a well-established algorithm with a rich history and is widely available in various machine learning libraries and software packages.

Disadvantages of Linear Regression

- Linear regression assumes a linear relationship between the dependent and independent variables. If the relationship is not linear, the model may not perform well.
- Linear regression is sensitive to multicollinearity, which occurs when there is a high correlation between independent variables. Multicollinearity can inflate the variance of the coefficients and lead to unstable model predictions.
- Linear regression assumes that the features are already in a suitable form for the model. Feature engineering may be required to transform features into a format that can be effectively used by the model.
- Linear regression is susceptible to both overfitting and underfitting. Overfitting occurs when the model learns the training data too well and fails to generalize to unseen data. Underfitting occurs when the model is too simple to capture the underlying relationships in the data.
- Linear regression provides limited explanatory power for complex relationships between variables. More advanced machine learning techniques may be necessary for deeper insights

Conclusion

Linear regression is a fundamental machine learning algorithm that has been widely used for many years due to its simplicity, interpretability, and efficiency. It is a valuable tool for understanding relationships between variables and making predictions in a variety of applications.

However, it is important to be aware of its limitations, such as its assumption of linearity and sensitivity to multicollinearity. When these limitations are carefully considered, linear regression can be a powerful tool for data analysis and prediction

Linear Regression – Frequently Asked Questions (FAQs)

What does linear regression mean in simple?

Linear regression is a supervised machine learning algorithm that predicts a continuous target variable based on one or more independent variables. It assumes a linear relationship between the dependent and independent variables and uses a linear equation to model this relationship.

Why do we use linear regression?

Linear regression is commonly used for:

- *Predicting numerical values based on input features*
- *Forecasting future trends based on historical data*
- *Identifying correlations between variables*
- *Understanding the impact of different factors on a particular outcome*

How to use linear regression?

Use linear regression by fitting a line to predict the relationship between variables, understanding coefficients, and making predictions based on input values for informed decision-making.

Why is it called linear regression?

Linear regression is named for its use of a linear equation to model the relationship between variables, representing a straight line fit to the data points.

What is linear regression examples?

Predicting house prices based on square footage, estimating exam scores from study hours, and forecasting sales using advertising spending are examples of linear regression applications.

Gradient Descent in Linear Regression

We know that in any machine learning project our main aim relies on how good our project accuracy is or how much our model prediction differs from the actual data point. Based on the difference between model prediction and actual data points we try to find the parameters of the model which give better accuracy on our dataset\, In order to find these parameters we apply gradient descent on the cost function of the machine learning model.

What is Gradient Descent

Gradient Descent is an iterative optimization algorithm that tries to find the optimum value (Minimum/Maximum) of an objective function. It is one of the most used optimization techniques in machine learning projects for updating the parameters of a model in order to minimize a cost function.

The main aim of gradient descent is to find the best parameters of a model which gives the highest accuracy on training as well as testing datasets. In gradient descent, The gradient is a vector that points in the direction of the steepest increase of the function at a specific point. Moving in the opposite direction of the gradient allows the algorithm to gradually descend towards lower values of the function, and eventually reaching to the minimum of the function.

Steps Required in Gradient Descent Algorithm

- **Step 1** we first initialize the parameters of the model randomly
- **Step 2** Compute the gradient of the cost function with respect to each parameter. It involves making partial differentiation of cost function with respect to the parameters.
- **Step 3** Update the parameters of the model by taking steps in the opposite direction of the model. Here we choose a hyperparameter learning rate which is denoted by alpha. It helps in deciding the step size of the gradient.
- **Step 4** Repeat steps 2 and 3 iteratively to get the best parameter for the defined model

Mathematics Behind Gradient Descent

In the Machine Learning Regression problem, our model targets to get the best-fit regression line to predict the value y based on the given input value (x). While training the model, the model calculates the cost function like Root Mean Squared error between the predicted value (pred) and true value (y). Our model targets to minimize this cost function .

To minimize this cost function, the model needs to have the best value of θ1 and θ2(for Univariate linear regression problem). Initially model selects θ1 and θ2 values randomly and then iteratively update these value in order to minimize the cost function until it reaches the minimum. By the time model achieves the minimum cost function, it will have the best θ1 and θ2 values. Using these updated values of θ1 and θ2 in the hypothesis equation of linear equation, our model will predict the output value y.

How Does Gradient Descent Work

Gradient descent works by moving downward toward the pits or valleys in the graph to find the minimum value. This is achieved by taking the derivative of the cost function, as illustrated in the figure below. During each iteration, gradient descent step-downs the cost function in the direction of the steepest descent. By adjusting the parameters in this direction, it seeks to reach the minimum of the cost function and find the best-fit values for the parameters. The size of each step is determined by parameter **α** known as **Learning Rate**.

How To Choose Learning Rate

The choice of correct learning rate is very important as it ensures that Gradient Descent converges in a reasonable time. :

- If we choose **α to be very large**, Gradient Descent can overshoot the minimum. It may fail to converge or even diverge.

If we choose α to be very small, Gradient Descent will take small steps to reach local minima and will take a longer time to reach minima.

Advantages Of Gradient Descent

- **Flexibility:** Gradient Descent can be used with various cost functions and can handle non-linear regression problems.
- **Scalability:** Gradient Descent is scalable to large datasets since it updates the parameters for each training example one at a time.
- **Convergence:** Gradient Descent can converge to the global minimum of the cost function, provided that the learning rate is set appropriately.

Disadvantages Of Gradient Descent

- **Sensitivity to Learning Rate**: The choice of learning rate can be critical in Gradient Descent since using a high learning rate can cause the algorithm to overshoot the minimum, while a low learning rate can make the algorithm converge slowly.
- **Slow Convergence:** Gradient Descent may require more iterations to converge to the minimum since it updates the parameters for each training example one at a time.
- **Local Minima:** Gradient Descent can get stuck in local minima if the cost function has multiple local minima.
- **Noisy updates:** The updates in Gradient Descent are noisy and have a high variance, which can make the optimization process less stable and lead to oscillations around the minimum.

Overall, Gradient Descent is a useful optimization algorithm for linear regression, but it has some limitations and requires careful tuning of the learning rate to ensure convergence.

Normal Equation in Linear Regression

Linear regression is a popular method for understanding how different factors (independent variables) affect an outcome (dependent variable. At its core, linear regression aims to find the best-fitting line that minimizes the error between observed data points and predicted values. One efficient method to achieve this is through the use of the normal equation. In this article, we will understand in-depth into the details of the normal equation, its mathematical derivation, implementation, and comparison with other optimization methods like gradient descent.

Understanding Normal Equation in Linear Regression

The normal equation is a mathematical formula that provides a straightforward way to calculate the coefficients (β\betaβ) in linear regression. Instead of using trial-and-error or iterative methods, the normal equation allows us to find the best coefficients directly.

The Normal Equation vs Gradient Descent

These are the two primary methods for estimating the coefficients (parameters) in Linear Regression Model. Each method has its unique advantages and considerations, making them suitable for different scenarios. Let's see a detailed comparison between these two approaches. ***The normal equation provides an analytical solution, computing the optimal parameters in a single step. Gradient descent uses an iterative approach, updating the parameters until convergence.***

The Normal Equation provides a closed-form solution to linear regression, allowing for the computation of optimal coefficients in one step.

- This method is efficient for small to medium-sized datasets because it relies on straightforward matrix operations.
- However, its dependence on matrix inversion can become computationally expensive with large datasets.
- Additionally, the Normal Equation does not require hyperparameter tuning, making its implementation simpler.

In contrast, **Gradient Descent** is an iterative optimization algorithm that adjusts the coefficients incrementally based on the gradient of the cost function.

- This method is particularly effective for large datasets, as it can process data points one at a time or in mini-batches, thereby reducing memory requirements.
- While Gradient Descent may take longer to converge compared to the Normal Equation, its performance can be significantly influenced by the choice of learning rate, which necessitates some degree of hyperparameter tuning.
- Furthermore, there are several variants of Gradient Descent, including stochastic and mini-batch methods, that can improve convergence speed and model generalization.

In summary, the choice between the normal equation and gradient descent often depends on the characteristics of the dataset and the problem at hand:

- ***Use the Normal Equation***:
    - *When working with smaller datasets or when a quick solution is needed without iterative tuning.*
    - *When simplicity and directness are prioritized.*
- ***Use Gradient Descent***:
    - *When handling large datasets or when the feature set is extensive, which could make matrix inversion computationally prohibitive.*
    - *When a more flexible approach is needed, especially if the model complexity increases or if you're working with online learning scenarios.*

Both methods are fundamental to understanding linear regression, and knowing when to use each can significantly impact the efficiency and effectiveness of your machine learning solutions.

Conclusion

The normal equation is a powerful tool for solving linear regression problems analytically. It provides a straightforward method to find the optimal parameters that minimize the sum of squared errors. However, its computational complexity makes it less suitable for large datasets. In such cases, gradient descent or other iterative methods are more efficient.

Normal Equation in Linear Regression – FAQs

What is the normal equation in Machine Learning?

The Normal Equation is a mathematical formula used to compute optimal parameters in linear regression by minimizing the sum of squared errors.

What Are the Advantages and Disadvantages of Using the Normal Equation?

Advantages: *Provides a direct solution; no iterative process needed.*
Disadvantages: *Computationally expensive for large datasets due to matrix inversion; less flexible for non-linear models.*

What Are the Practical Considerations for Implementing the Normal Equation in Python?

Ensure data is preprocessed and scaled; handle singular matrices; use efficient libraries like NumPy for matrix operations to enhance performance.

Generalized Linear Models

Generalized Linear Models (GLMs) are a class of regression models that can be used to model a wide range of relationships between a response variable and one or more predictor variables. Unlike traditional linear regression models, which assume a linear relationship between the response and predictor variables, GLMs allow for more flexible, non-linear relationships by using a different underlying statistical distribution.

Some of the features of GLMs include:

1. Flexibility: GLMs can model a wide range of relationships between the response and predictor variables, including linear, logistic, Poisson, and exponential relationships.
2. Model interpretability: GLMs provide a clear interpretation of the relationship between the response and predictor variables, as well as the effect of each predictor on the response.
3. Robustness: GLMs can be robust to outliers and other anomalies in the data, as they allow for non-normal distributions of the response variable.
4. Scalability: GLMs can be used for large datasets and complex models, as they have efficient algorithms for model fitting and prediction.
5. Ease of use: GLMs are relatively easy to understand and use, especially compared to more complex models such as neural networks or decision trees.
6. Hypothesis testing: GLMs allow for hypothesis testing and statistical inference, which can be useful in many applications where it's important to understand the significance of relationships between variables.
7. Regularization: GLMs can be regularized to reduce overfitting and improve model performance, using techniques such as Lasso, Ridge, or Elastic Net regression.
8. Model comparison: GLMs can be compared using information criteria such as AIC or BIC, which can help to choose the best model among a set of alternatives

Some of the disadvantages of GLMs include:

- Assumptions: GLMs make certain assumptions about the distribution of the response variable, and these assumptions may not always hold.
- Model specification: Specifying the correct underlying statistical distribution for a GLM can be challenging, and incorrect specification can result in biased or incorrect predictions.
- Overfitting: Like other regression models, GLMs can be prone to overfitting if the model is too complex or has too many predictor variables.
- Overall, GLMs are a powerful and flexible tool for modeling relationships between response and predictor variables, and are widely used in many fields, including finance, marketing, and epidemiology. If you're interested in learning

more about GLMs, you might consider reading an introductory textbook on regression analysis, such as “An Introduction to Generalized Linear Models” by Annette J. Dobson and Annette J. Barnett.

- Limited flexibility: While GLMs are more flexible than traditional linear regression models, they may still not be able to capture more complex relationships between variables, such as interactions or non-linear effects.
- Data requirements: GLMs require a sufficient amount of data to estimate model parameters and make accurate predictions, and may not perform well with small or imbalanced datasets.
- Model assumptions: GLMs rely on certain assumptions about the distribution of the response variable and the relationship between the response and predictor variables, and violation of these assumptions can lead to biased or incorrect predictions.

Locally weighted Linear Regression

- Locally weighted linear regression is a supervised learning algorithm.
- It is a non-parametric algorithm.
- There exists No training phase. All the work is done during the testing phase/while making predictions.
- The dataset must always be available for predictions.
- Locally weighted regression methods are a generalization of k-Nearest Neighbour.
- In Locally weighted regression an explicit local approximation is constructed from the target function for each query instance.
- The local approximation is based on the target function of the form like constant, linear, or quadratic functions localized kernel functions.