AIML

Nitin Arvind Shelke

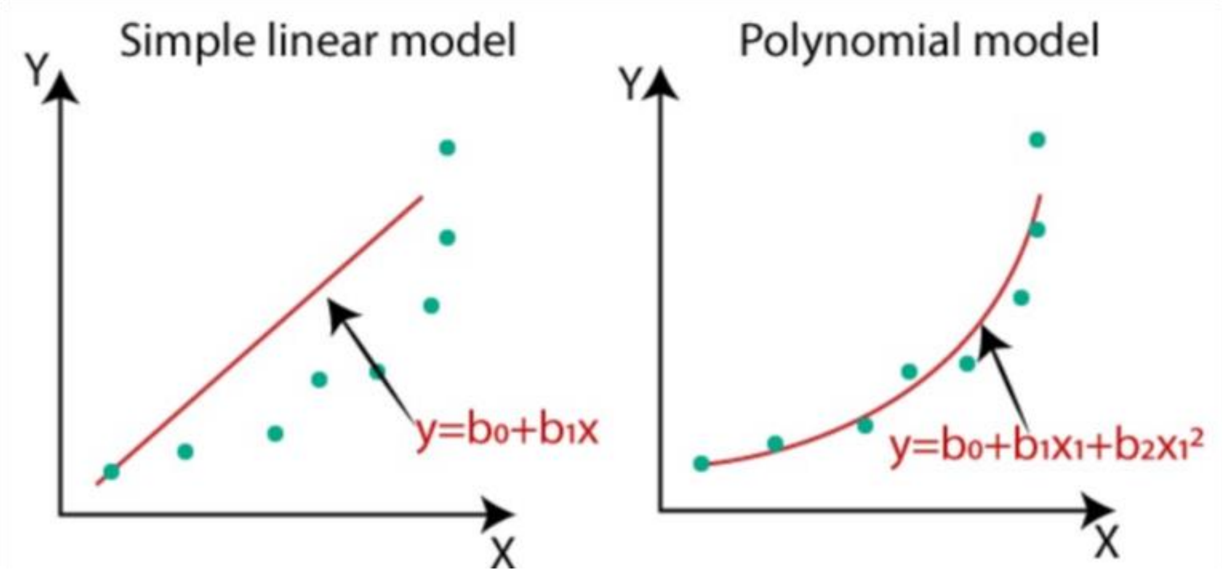
Polynomial Regression

- Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial. The Polynomial Regression equation is given below:
- Y=a0+a1x+a2x^2+a3x^3+a4x^4+a5x^5
- Polynomial Regression is a type of regression which models the non-linear dataset using a linear model.
- It is also called the special case of Multiple Linear Regression in ML. Because we add some polynomial terms to the Multiple Linear regression equation to convert it into Polynomial Regression.

Need for Polynomial Regression:

If we apply a linear model on a linear dataset, then it provides us a good result as we have seen in Simple Linear Regression, but if we apply the same model without any modification on a non-linear dataset, then it will produce a drastic output. Due to which loss function will increase, the error rate will be high, and accuracy will be decreased.

So for such cases, where data points are arranged in a non-linear fashion, we need the Polynomial Regression model. We can understand it in a better way using the below comparison diagram of the linear dataset and non-linear dataset.



Ridge Regression vs Lasso Regression

- Ridge and Lasso Regression are two popular techniques in AIML used for regularizing linear models to avoid overfitting and improve predictive performance. Both methods add a penalty term to the model’s cost function to constrain the coefficients, but they differ in how they apply this penalty.

- Ridge Regression, also known as L2 regularization, adds the squared magnitude of the coefficients as a penalty. On the other hand, Lasso Regression, or L1 regularization, introduces a penalty based on the absolute value of the coefficients.

Ridge Regression

- Ridge regression, also known as L2 regularization, is a technique used in linear regression to prevent overfitting by adding a penalty term to the loss function. This penalty is proportional to the square of the magnitude of the coefficients (weights).
- Objective Function=MSE+ Lambda*Sum of Square of Coefficients

Ridge Regression

Ridge adds an L_2-norm penalty to the loss function to shrink the coefficients but does not enforce sparsity.

$$\text{Minimize: } \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

- **MSE Term:** $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- **Penalty Term:** $\lambda \sum_{j=1}^{p} \beta_j^2$ (squared values of coefficients).
- λ: Regularization strength. Higher λ shrinks coefficients more but does not set any to 0.

Lasso Regression

- Lasso regression, also known as L1 regularization, is a linear regression technique that adds a penalty to the loss function to prevent overfitting. This penalty is based on the absolute values of the coefficients.

```
Objective Function=MSE + Lambda*Sum of Absolute coefficients
```

- Lasso regression is a version of linear regression that includes a penalty equal to the absolute value of the coefficient magnitude. By encouraging sparsity, this L1 regularization term reduces overfitting and helps some coefficients to be absolutely zero, hence facilitating feature selection.

Lasso Regression

Lasso adds an L_1-norm penalty to the loss function to encourage sparsity in the coefficients.

$$\text{Minimize: } \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

- **MSE Term:** $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- **Penalty Term:** $\lambda\sum_{j=1}^{p}|\beta_j|$ (absolute values of coefficients).
- λ: Regularization strength. Higher λ shrinks coefficients more, potentially setting some to 0.

Lasso Vs Ridge

Characteristic	Ridge Regression	Lasso Regression
Penalty Type	L2 (squared magnitude of coefficients)	L1 (absolute magnitude of coefficients)
Coefficient Shrinkage	Shrinks coefficients but doesn't force them to zero	Can shrink some coefficients to exactly zero
Feature Selection	Does not perform feature selection	Performs automatic feature selection by eliminating irrelevant or redundant features
Model Complexity	Tends to include all features in the model	Can simplify the model by excluding some features
Impact on Prediction	Tends to handle multicollinearity well	Can simplify the model which might improve prediction for high-dimensional data
Best for	Use when all features are important, and you want to regularize the coefficients without removing features.	Use when feature selection is important, and you expect some features to be irrelevant.

Elastic Net

- Elastic Net is a regularized regression technique that combines both Lasso and Ridge penalties. It is particularly useful when there are correlated features or when neither Lasso nor Ridge alone provides optimal results.

Elastic Net

Objective Function:

$$\text{Minimize:} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2$$

- $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ Mean Squared Error (MSE) term.
- λ_1: Coefficient for the L_1-norm (Lasso penalty).
- λ_2: Coefficient for the L_2-norm (Ridge penalty).

Gradient Descent for Regression

- Gradient Descent is an optimization algorithm used to minimize the cost function of a model. For regression problems, the goal is to find the values of coefficients (βj) that minimize the cost function (e.g., Mean Squared Error).

Regression Performance Metrics

- Mean Absolute Error(MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{y}_i\right|$$

Regression Performance Metrics

- Mean Squared Error(MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2$$

Regression Performance Metrics

- Root Mean Squared Error(RMSE)

$$\text{RMSE} = \sqrt{\text{MSE}}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)}$$

Regression Performance Metrics

- R2Score or The coefficient of determination

$$R^2 = 1 - \frac{\sum_{1=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{1=1}^{N} (y_i - \bar{y})^2}$$