

Non Linear Regression

It is a statistical technique used to model complex relationship between variables that cannot be adequately described by a straight line or linear equation.

A Non linear regression can be expressed as :

$$Y = f(X, \beta) + \epsilon$$

Where,

$f(X, \beta)$ Regression function

X → Vector independent variables
Used to predict dependent variable.

β → Vector of parameters that model aims to estimate

↳ These parameters determine the shape and characteristics of the regression function.

ϵ → Error term

Types:

* Parametric → Assume that variables modeled using mathematical function.

* Non-parametric → does not assume modeled using mathematical function.

Polynomial Regression

Powerful non linear technique.

Capture more complex relationships between the independent and dependent variables.

Uses higher order polynomial functions

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n$$

↑ ↓ ↓ ↓
 Dependent Variable Independent Variable Parameters of the model

n — degree of polynomial

Exponential Regression

models data following an exponential growth or decay pattern.

particularly useful for analyzing phenomena that exhibit rapid growth or decline such as population growth, spread of infectious diseases.

$y = \alpha e^{(\beta x)}$

α and β — parameters of the model.

Logistic Regression.

- Binary Classification
- Probability Estimation.
- Non Linear relationships

Advantages of Non Linear regression.

- * **Flexibility**
Capture Complex, non-linear relationships between variables.
- * **Improved Accuracy**
It can fit data better than linear models, leading to more accurate predictions.
- * **Representation of Real-World Phenomena**
Many real-world processes exhibit non-linear behavior, which non-linear models are better equipped to capture.

Applications

- * Medical Research
- * Business Forecasting
- * Scientific Research
- * Urban Planning

Linearization Transforms

It is used to approximate non-linear models by converting them into a linear form.

So that linear regression techniques can be applied.

Common Techniques:

Logarithmic Transformation:

Applied when the relationship is exponential.

$$\text{Eg: } y = a \cdot e^{bx}$$

$$\log(y) = \log(a) + bx.$$

Reciprocal Transformation:

When the relationship is hyperbolic.

$$\text{Eg: } y = \frac{a}{x+b} \quad \text{If can be}$$

linearized by inverting both sides.

Polynomial Transformation:

Applied when the relationship is curvilinear.

To assess whether a model is non-linear,

various diagnostic methods are used:

- * Scatter plot
- * Goodness of fit tests
- * Partial residual plots

Example :

Suppose non-linear equation is of the form $y = a \cdot e^{bx}$

after taking \ln

$$\ln(y) = \ln(a) + bx.$$

$$Y = \ln(y)$$

$$\beta_0 = \ln(a) \quad (\text{Constant})$$

$$\beta_1 = b \quad (\text{slope})$$

Equation now becomes

$$Y = \beta_0 + \beta_1 x$$

Estimate β_0 and β_1 , and then back-transform to find the original parameters.

$$a = e^{\beta_0} = (Y)_{\text{ad}}$$

$$b = \beta_1$$

Assume

x	y
1	2.7
2	7.4
3	20.1
4	54.6

first apply natural log to the y va

x	ln(y)
1	0.993
2	2.001
3	3.004
4	4.000

after performing linear regression on the transformed data $(x, \ln(y))$

The estimated linear regression equation is

$$\ln(y) = \beta_0 + \beta_1 x$$

by least square

$$\boxed{\ln(y) = 0.5 + 0.8x}$$

from the linear regression Output

$$\beta = 0.5 \rightarrow a = e^{0.5} = 1.648$$

$$\beta = 0.8 \rightarrow b = 0.8$$

Thus, the estimated parameters for the original non-linear model are:

$$y = 1.648 \cdot e^{0.8x}$$

Iterative Procedures for Non-Linear Least Squares (NLS)

The parameters can not be estimated using direct algebraic solutions like linear regression.

Iterative procedures are used to find the best-fitting parameters.

These procedures aim to minimize the sum of squared residuals (RSS)

i.e. $S(\beta)$

$$S(\beta) = \sum_{i=1}^n (y_i - f(x_i, \beta))^2$$

y_i - Observed Value

$f(x_i, \beta)$ - model predicted value using β

$S(\beta)$ = Sum of squared residuals.

Common iterative procedures for NL

1. Grid Search

Brute-force method that involves searching over a predefined grid of possible parameter values.

At each point on the grid the value of sum of squared residuals $S(\beta)$ is calculated and the parameters that minimize $S(\beta)$ are chosen.

Steps:

1. Define a grid of possible values for each parameter.

2. Compute error function $S(\beta)$ for each combination of parameter values.

3. Choose the combination of parameters that minimizes the error.

Adv:

* Simple and straightforward to implement.

* No derivatives (or) complex mathematical operations are required.

Disadv

Computationally expensive
Grid resolution affect Accuracy.

2. Newton - Raphson Method

It is an iterative procedure that uses second-order information (Hessian matrix of second derivatives) to refine parameter estimates.

It is fast and converges rapidly near the solution, but it requires the calculation of derivatives.

Steps:

1. Start with an initial guess for parameters vector β_0

2. Update β_0 by using

$$\beta_{\text{new}} = \beta_{\text{old}} - H^{-1}(\beta_{\text{old}}) \cdot \nabla S(\beta_{\text{old}})$$

$H(\beta_{\text{old}})$ - Hessian matrix of second derivatives of $S(\beta)$

$\nabla S(\beta_{\text{old}})$ is the gradient (vector of first derivatives) of $S(\beta)$.

3. Repeat the process until the changes in β converge.

Adv:

- * Very fast convergence if the starting point is near the solution
- * uses second order information for precise updates.

Disadv:

- * Requires the computation of second derivatives, complex and computationally expensive
- * May diverge if the initial guess is far from true solution.
- * Hessian matrix may not be invertible or stable, leading to numerical issues.

3. Steepest Descent (Gradient Descent)

This method is a first-order optimization technique that updates the parameters in the direction of the negative gradient of the objective function.

It is widely used due to its simplicity.

Steps:

1. Start with an initial guess for the parameter vector β_0 .

2. Update the parameters by moving in the direction opposite to gradient of $S(\beta)$.

$$\beta_{\text{new}} = \beta_{\text{old}} - \alpha \cdot \nabla S(\beta_{\text{old}})$$

Where :

α is the learning rate (Step size)

$\nabla S(\beta_{\text{old}})$ is the gradient of Objective function $S(\beta)$.

3. Continue updating until the change in the objective function (or) parameters is sufficiently small.

Adv: Adv:

- * Simple and easy to implement
- * Requires only first-order derivatives (gradient), making it computationally cheaper than Newton-Raphson.
- * Can be used for large datasets and models with many parameters.

Disadv:

- * Convergence can be slow
- * May get stuck in local minima.
- * Sensitive to the choice of the learning rate α .

4. Marquardt's Method (Levenberg - Marquardt Algorithm)

It is a hybrid technique that combines aspects of both Newton-Raphson and steepest descent.

It switches between these two methods based on the current approximation to the parameters, making it more robust in practice.

Steps:

1. Start with an initial guess for the parameter vector β_0 .
2. At each iteration, update the parameters using the following

$$\boxed{\beta_{\text{new}} = \beta_{\text{old}} - (J^T J + \lambda I)^{-1} J^T r}$$

J - Jacobian matrix [matrix of first

derivatives of the residuals]

r - vector of residuals,

λ - damping parameter that controls between gradient and Newton's method

I - identity matrix.

3. When λ is large, algorithm behaves like steepest descent.

4. λ is small, algorithm behaves like Newton-Raphson

5. Adjust λ dynamically during the iterations based on the success of the step.

Adv:

- * Works well for a wide variety of non-linear problems.
- * Reduces the likelihood of divergence.
- * Combines the robustness of gradient descent with speed of Newton-Raphson.

Disadv:

- * More computationally demanding than gradient descent.
- * Requires tuning the damping parameter λ which can be tricky.

$$-3 + (x_1)^2 + 1x_1^2 + 0.2 = 0$$

gradient search - 1st
Newton-Raphson - 2nd
(eg. 2nd grad. has old x₁)

Semiparametric Regression Models

These are statistical models that combine both parametric and nonparametric components.

It allows for more flexibility than fully parametric models but retain more structure than nonparametric models.

Example

Predicting income (y) based on age (x_1) and education level (x_2).

Income increases linearly with age (parametric part), but the effect of education might be more complex and not strictly linear (nonparametric part).

A semiparametric model could look like

$$y = \beta_0 + \beta_1 x_1 + f(x_2) + \epsilon$$

$\beta_1 x_1$ - linear relationship

$f(x_2)$ - nonparametric function (This function is flexible and take any shape.)

Adv:

- * flexibility to model complex relationships without fully sacrificing interpretability.
- * Avoid making strict assumptions about the entire model form.

Additive Regression Models

These are a specific type of semiparametric model where the relationship between the response variable and predictors modeled as sum of functions, each depending on a single predictor variable.

Additive model allows non-linear relationship between each predictor and the response.

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + e$$

β_0 Intercept

$f_1, f_2, f_3, \dots, f_p$ are smooth, possibly non-linear functions of the predictors x_1, x_2, \dots, x_p .

Additive models are typically fitted using techniques such as smoothing splines (or) penalized regression splines.

Example

predicting house prices (y) based on square footage (x_1), Distance from the City Center (x_2).
 additive model

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \epsilon$$

Capture the non-linear effect of Square footage on house price

→ models the Nonlinear effect of distance on price

Benefit:

Each predictor's effect can be interpreted separately

This allows the response variable to depend on each predictor through an independent, smooth function. The total response is the sum of these individual effects, without assuming any interaction between predictors.

Non Parametric Regression Methods

It do not assume a predefined functional form for the relationship between the response variable and predictors.

These methods are entirely data driven and rely on the structure of the data to infer the relationship.

particularly useful when there is little prior knowledge about the form of the underlying relationship (or) when the relationship is too complex to be captured by parametric models.

Common Nonparametric methods

* Kernel regression: Uses a weighted average of nearby data points, with weights determined by kernel function.

* Smoothing Splines: Fit smooth curves through the data by balancing data fit and smoothness.

* Local regression: (LOESS (LOWELL)): performs regression locally at each point using nearby data points.

Example

predict bike rental demand (y)
based on temperature (x).

We don't know the form of
relationship between temperature and
demand.

Use kernel regression

Bike rental demand Vs temperature

- * At low temperatures, demand is low
- * As temperature rises, demand increases but once it becomes too hot, demand drops off.

Benefits

It captures complex relationships directly from the data, without requiring us to specify the form of the relationship in advance.