1. Introduction to Regression

1.1 What is Regression?

1.1.1 Definition and Purpose

Regression is a statistical method used to understand the relationship between a dependent variable (target) and one or more independent variables (features). The main purpose of regression is to model this relationship in order to predict the dependent variable based on the values of the independent variables.

1.1.2 Types of Regression

1. **Linear Regression**:
    - Models the relationship between the dependent and independent variables as a straight line.
    - Equation: $Y = \beta 0 + \beta 1 X + \epsilon$
    - Used when the relationship between variables is assumed to be linear.

2. **Polynomial Regression**:
    - Models the relationship as an nth degree polynomial.
    - Equation: $Y = \beta 0 + \beta 1 X + \beta 2 X2 + \cdots + \beta n Xn + \epsilon$
    - Useful when the data shows a nonlinear relationship.

3. **Logistic Regression**:
    - Used for binary classification problems.
    - Models the probability that a given input point belongs to a certain class.
    - Equation: $P(Y = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$

4. **Ridge Regression**:
    - A type of linear regression that includes a regularization term to prevent overfitting.
    - Equation: $Y = \beta_0 + \beta_1 X + \epsilon + \lambda \sum_{i=1}^{n} \beta_i^2$

5. **Lasso Regression**:
    - Similar to ridge regression but uses L1 regularization, which can shrink some coefficients to zero, effectively performing variable selection.
    - Equation: $Y = \beta_0 + \beta_1 X + \epsilon + \lambda \sum_{i=1}^{n} |\beta_i|$

1.2 Use Cases of Regression

1.2.1 Predicting Continuous Outcomes

Regression is primarily used to predict continuous outcomes. For example:

- **House Prices**: Predicting the price of a house based on features like location, size, number of rooms, etc.
- **Sales Forecasting**: Estimating future sales based on historical sales data and other variables such as advertising spend.

1.2.2 Applications in Various Domains

- **Finance**:
    - Predicting stock prices, bond yields, and other financial metrics.
    - Risk assessment and credit scoring.
- **Healthcare**:
    - Predicting patient outcomes based on medical history and current health indicators.
    - Estimating the spread of diseases.
- **Marketing**:
    - Analyzing customer data to predict future purchasing behavior.
    - Effectiveness of marketing campaigns.
- **Economics**:
    - Forecasting economic indicators such as GDP, inflation rates, and employment levels.
- **Engineering**:
    - Modelling physical systems and predicting system behaviors under different conditions.

Regression analysis is a versatile tool that finds applications across a wide range of fields, helping organizations make informed decisions based on data-driven insights.

2. Linear Regression

2.1 Understanding Linear Regression

2.1.1 Model Representation

Linear regression models the relationship between a dependent variable Y and one or more independent variables X. The simplest form is simple linear regression, which involves a single predictor

$$Y = \beta 0 + \beta 1X + \epsilon$$

- $\beta 0$ is the intercept.
- $\beta 1$ is the slope of the line.
- ϵ is the error term.

For multiple linear regression, the model includes multiple predictors

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn + \epsilon$$

2.1.2 Assumptions of Linear Regression

1. **Linearity**: The relationship between the dependent and independent variables is linear.
2. **Independence**: Observations are independent of each other.
3. **Homoscedasticity**: The residuals (errors) have constant variance at every level of X.
4. **Normality**: The residuals of the model are normally distributed.
5. **No Multicollinearity**: Independent variables are not highly correlated with each other.

2.2 Mathematical Formulation

2.2.1 Linear Equation

The linear equation in matrix form is: $Y=X\beta+\epsilon$

where:

- Y is the vector of observed values.
- X is the matrix of input features.
- β is the vector of coefficients.
- ϵ is the vector of errors.

2.2.2 Cost Function (Mean Squared Error)

The cost function used to measure the accuracy of the linear regression model is the Mean Squared Error (MSE)

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\beta(\mathbf{x}^{(i)}) - y^{(i)})^2$$

where:

- m is the number of training examples.
- $h_\beta(\mathbf{x}^{(i)})$ is the predicted value.
- $y^{(i)}$ is the actual value.

2.2.3 Gradient Descent Optimization

Gradient descent is an optimization algorithm used to minimize the cost function. The update rule for the parameters

$$\beta_j := \beta_j - \alpha \frac{\partial J(\beta)}{\partial \beta_j}$$

where:

- α is the learning rate.
- $\frac{\partial J(\beta)}{\partial \beta_j}$ is the partial derivative of the cost function with respect to β_j.

---

3. Polynomial Regression

3.1 Understanding Polynomial Regression

3.1.1 Difference from Linear Regression

The key difference between linear and polynomial regression is in the type of relationship they model between the independent and dependent variables. While linear regression models a linear relationship (a straight line), polynomial regression can model a nonlinear relationship (a curve). This is achieved by including polynomial terms (squared, cubed, etc.) of the independent variables in the model.

3.1.2 When to Use Polynomial Regression

Polynomial regression is used when the data shows a curvilinear relationship between the independent and dependent variables. If a linear model is insufficient to capture the trends in the data, polynomial regression can provide a better fit.

3.2 Mathematical Formulation

3.2.1 Polynomial Equation

A polynomial regression model represents the relationship between the dependent variable Y and the independent variable X as an nth degree polynomial

$$Y = \beta0 + \beta1X + \beta2X2 + \beta3X3 + \cdots + \beta nXn + \epsilon$$

- β0, β1, β2, β3, …, βn are the coefficients.
- ϵ is the error term.

3.2.2 Cost Function and Optimization

The cost function for polynomial regression is the same as for linear regression, typically the Mean Squared Error (MSE):

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\beta(\mathbf{x}^{(i)}) - y^{(i)})^2$$

Gradient descent can be used to minimize this cost function, updating the coefficients to reduce the error.

---

4. Regularization Techniques

4.1 Introduction to Regularization

4.1.1 Purpose of Regularization

Regularization is a technique used to improve the generalizability of a machine learning model by adding a penalty for large coefficients in the model. This helps to prevent overfitting, where the model performs well on the training data but poorly on unseen test data. Regularization can also help with feature selection, making the model simpler and more interpretable.

4.1.2 Overfitting and Underfitting

- **Overfitting**: Occurs when a model is too complex, capturing noise and details in the training data that do not generalize to new data. Regularization can help reduce overfitting by penalizing large coefficients.
- **Underfitting**: Occurs when a model is too simple to capture the underlying structure of the data. It results in poor performance on both the training and test data. Regularization techniques need to be balanced to avoid underfitting.

4.2 Ridge Regression (L2 Regularization)

4.2.1 Mathematical Formulation

Ridge regression adds an L2 penalty to the cost function, which is the sum of the squares of the coefficients

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\beta(\mathbf{x}^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \beta_j^2$$

where λ is the regularization parameter that controls the strength of the penalty.

4.3 Lasso Regression (L1 Regularization)

4.3.1 Mathematical Formulation

Lasso regression adds an L1 penalty to the cost function, which is the sum of the absolute values of the coefficients

$$J(\beta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\beta(\mathbf{x}^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

where λ is the regularization parameter that controls the strength of the penalty.

---

5. Model Evaluation and Metrics

5.1 Evaluation Metrics for Regression

5.1.1 Mean Absolute Error (MAE)

The Mean Absolute Error is the average of the absolute differences between predicted and actual values

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

5.1.2 Mean Squared Error (MSE)

The Mean Squared Error is the average of the squared differences between predicted and actual values

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

5.1.3 Root Mean Squared Error (RMSE)

The Root Mean Squared Error is the square root of the average of the squared differences between predicted and actual values

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

5.1.4 R-squared (Coefficient of Determination)

The R-squared value measures the proportion of the variance in the dependent variable that is predictable from the independent variables

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

$\bar{y}$ is the mean of the actual values.