Regression in Machine Learning

1. Overview & Definition

Regression is a fundamental supervised machine learning technique used to predict continuous numerical values by learning relationships between input features (X) and a continuous target variable (Y). Unlike classification (which predicts discrete labels or categories), regression estimates measurable outcomes , making it essential for forecasting , trend analysis, financial planning, and risk assessment .

2. Visual Overview of Regression Techniques

The diagram below summarizes the primary regression models and categories used in machine learning workflows :

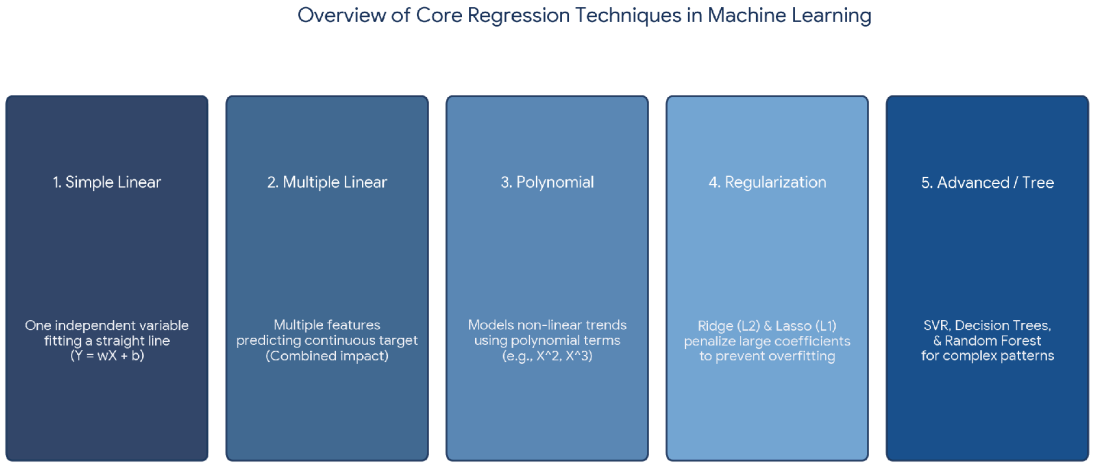

Figure 1: Core Regression Models and Techniques in Machine Learning

3. Major Types of Regression Techniques

3.1 Simple Linear Regression

Models the relationship between a single independent input variable and a continuous dependent target by fitting a straight line that minimizes the sum of squared errors .

> **Linear Equation:** $Y = w * X + b$
> *where w represents the weight/slope and b represents the bias/intercept.*

- **Advantages:** Highly interpretable and straightforward to compute .
- **Limitations:** Fails to capture complex, non-linear relationships .

3.2 Multiple Linear Regression

Extends simple linear regression by incorporating multiple independent predictor variables to predict a continuous outcome . Each predictor is assigned a coefficient reflecting its independent impact .

> **Multiple Linear Equation:** $Y = b + w_1*X_1 + w_2*X_2 + ... + w_n*X_n$

- **Advantages:** Captures the combined influence of multiple factors simultaneously .
- **Limitations:** Model performance degrades in the presence of multicollinearity (high correlation among features) .

3.3 Polynomial Regression

Models non-linear relationships by transforming input features into higher-degree polynomial terms (e.g., X^2, X^3) . Although the relationship is non-linear in features, it remains linear in its parameters, allowing standard linear optimization techniques .

> **Polynomial Equation:** $Y = b + w_1*X + w_2*X^2 + ... + w_n*X^n$

3.4 Ridge and Lasso Regularization

Regularized linear models that add penalty terms to limit large coefficient magnitudes and prevent overfitting :

- **Ridge Regression (L2):** Uses L2 penalty regularization to shrink coefficients smoothly .
- **Lasso Regression (L1):** Uses L1 penalty regularization, which can shrink certain coefficients completely to zero, performing automatic feature selection .

3.5 Advanced Regression Models

- **Support Vector Regression (SVR):** Applies Support Vector Machines to regression by fitting a function within a defined margin (epsilon-tube) and using kernel tricks for non-linear patterns .
- **Decision Tree Regression:** Recursively partitions feature space into hierarchical branches to predict continuous target values .
- **Random Forest Regression:** An ensemble bagging method that trains multiple decision trees and averages their predictions to reduce overfitting and boost robustness .

4. Core Summary Table: Regression Evaluation Metrics

Evaluation Metric	Description	Interpretation
Mean Absolute Error (MAE)	Average absolute difference between predicted and actual values	Robust to outliers; expressed in target units
Mean Squared Error (MSE)	Average squared difference between predicted and actual values	Penalizes large errors heavily due to squaring

Root Mean Squared Error (RMSE)	Square root of the mean squared error	Interpretable in the same units as the target variable
R-Squared (R^2 Score)	Proportion of variance in the dependent variable explained by predictors	Ranges from 0 to 1 (higher values indicate better fit)
Huber Loss	Hybrid loss function combining MAE and MSE characteristics	Provides robustness against outliers while maintaining smooth gradients

5. Common Real-World Applications

- **Price Prediction:** Estimating housing or vehicle market values based on attributes.
- **Trend Forecasting:** Predicting future sales , inventory demand, and financial market indexes.
- **Risk Analysis:** Assessing credit risk scores and patient health deterioration probabilities.