# Introduction to Regression in Machine Learning

## 1. What is Regression?

Regression is a type of supervised learning used to predict continuous, numerical values based on input features. Examples include predicting house prices, stock market trends, or a person's age.

### Key Idea:

• Regression → "How many?" (continuous output)
• Classification → "Which class?" (categorical output)

| Feature | Regression | Classification |
|---------|-----------|----------------|
| Output | Continuous value | Discrete label |
| Question | How many? | Which class? |
| Example | Predicting car price | Email spam vs not spam |

## 2. Types of Regression

### 2.1 Linear Regression

Models a straight-line relationship between one input (independent variable) and one output (dependent variable).

### 2.2 Multiple Linear Regression (MLR)

Handles two or more independent variables to predict a continuous dependent variable.
Advantages: handles multiple predictors, easy interpretation, scalable for large datasets.
Limitations: assumes linearity, sensitive to outliers, affected by multicollinearity.

### 2.3 Polynomial Regression

Extends linear regression by fitting curved (non-linear) relationships.
Advantages: captures non-linear growth, flexible and accurate.
Limitations: risk of overfitting, harder interpretation, sensitive to outliers.

### 2.4 Other Types

• Logistic Regression
• Ordinal Regression
• Multinomial Logistic Regression

## 3. Regression Metrics

| Metric | Description | Example |
|--------|-------------|---------|
| MAE | Mean Absolute Error – average absolute difference between predictions and actual values | MAE=5 → off |
| MSE | Mean Squared Error – penalizes large errors more strongly | Large error in |
| RMSE | Root Mean Squared Error – square root of MSE, same units as data | RMSE=10 → |
| $R^2$ | Coefficient of Determination – proportion of variance explained | $R^2$=0.95 → ex |

## 4. Underfitting vs Overfitting

• Underfitting: model too simple, misses important patterns.
• Overfitting: model too complex, memorizes training data but fails on new data.
Prevention Methods: Train/test split, Cross-validation, Early stopping, Regularization, Reduce model complexity.


## 5. Case Study: Regression in Business

**Title:** Predicting Retail Store Sales
**Type:** Multiple Linear Regression
**Goal:** Predict weekly sales for inventory and workforce optimization.
**Data:** historical sales, store size, promotions, holidays, inventory.

Results: Store size and promotions were strongest predictors, helping managers plan inventory and staff efficiently.


## 6. Comparison of Regression Types

| Type | Relationship | Shape | Example | Best For |
|------|-------------|-------|---------|----------|
| Linear Regression | Straight line | Line | Years of Experience → Salary | Simple predictio |
| Multiple Linear Regression | Multi-dimensional | Plane | Age, Education, Experience → Salary | Multi-factor pre |
| Polynomial Regression | Non-linear | Curve | Time, Time² → Plant Growth | Trends & growt |


## References

1. builtin.com/data-science/regression-machine-learning
2. spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression
3. medium.com/@data-overload/understanding-polynomial-regression-a-powerful-tool-for-complex-relationships-d2394a898fd6
4. ibm.com/think/topics/classification-vs-regression
5. net-informations.com/ml/mla/poly.htm
6. medium.com/data-science/8-simple-techniques-to-prevent-overfitting-4d443da2ef7d

✍■ *By: Burhaan Cabdulaahi*