

Statistical Regression

Statistical regression refers to a set of statistical methods used to estimate the relationship between a dependent variable and one or more independent variables. It's commonly used for **prediction**, **forecasting**, and finding **causal relationships**.

◆ 1. Linear Regression

✓ Definition

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a **linear equation** to observed data.

💡 Use Case

- Predicting house prices based on square footage and location
- Estimating sales based on advertising budget
- Forecasting stock prices based on historical trends

❓ When to Use Linear Regression?

- When the relationship between variables is **linear**
- The dependent variable is **continuous**
- Residuals (errors) are **normally distributed** and **homoscedastic** (constant variance)
- No or minimal multicollinearity between independent variables

◆ 2. Logistic Regression

✓ Definition

Logistic regression is a statistical method used for **binary classification** problems. Instead of predicting a continuous value, it predicts the **probability** that a given input belongs to a particular category.

💡 Use Case

- Predicting whether a customer will churn (yes/no)
- Determining if an email is spam or not
- Classifying if a tumor is malignant or benign

❓ When to Use Logistic Regression?

- When the dependent variable is **binary** (0 or 1, Yes/No)

- Data is **linearly separable** (at least approximately)
- You need **probability estimates** for classification
- You want **interpretable coefficients**

Glad you liked it! Here's the same format applied to:

- **ML Models:** Decision Tree, Random Forest, SVM
- **DL Models:** CNNs, RNNs, Transformers

Machine Learning Models

1. Decision Tree

Definition

A decision tree is a tree-structured model used for classification and regression. It splits data into branches based on feature values, leading to decision outcomes.

Use Case

- Credit risk analysis (approve/decline loan)
- Customer segmentation
- Medical diagnosis

Example

A tree might split:

- Is income > \$50k?
 - Yes → Is credit score > 700?
 - Yes → Approve loan
 - No → Reject
 - No → Reject

When to Use?

- When interpretability is important
- For quick prototyping or baseline models
- When data contains **non-linear relationships**
- Works well with both **categorical and numerical data**

2. Random Forest

Definition

Random Forest is an **ensemble** of decision trees. It combines multiple trees to improve prediction accuracy and control overfitting by averaging or voting the output.

Use Case

- Predicting product ratings
- Fraud detection

- Image classification (basic)

? When to Use?

- When overfitting is a problem with a single decision tree
- When you need **high accuracy** over interpretability
- Works well with **large feature sets** and **missing data**

◆ 3. Support Vector Machine (SVM)

✓ Definition

SVM is a supervised learning algorithm that finds the **optimal hyperplane** to separate data into classes. It maximizes the **margin** between classes.

💡 Use Case

- Handwritten digit recognition
- Face detection
- Bioinformatics (classifying genes)

? When to Use?

- When data is **linearly (or non-linearly via kernel) separable**
- High-dimensional data (e.g. text classification)
- When you want a **robust margin** between classes
- Not ideal for large datasets (slow training)

Deep Learning Models

1. Convolutional Neural Networks (CNNs)

Definition

CNNs are deep learning models primarily used for **image and spatial data**. They use convolutional layers to detect patterns like edges, textures, shapes, etc.

Use Case

- Image classification (e.g. cats vs dogs)
- Object detection (YOLO, SSD)
- Medical imaging (tumor detection)

When to Use?

- When dealing with **image, video, or spatial data**
- Problems requiring **feature extraction** from local areas
- Tasks involving visual understanding or pattern detection

2. Recurrent Neural Networks (RNNs)

Definition

RNNs are neural networks designed for **sequential data**. They maintain memory of previous inputs through **recurrent connections**.

Use Case

- Text generation
- Sentiment analysis
- Time-series forecasting
- Speech recognition

When to Use?

- When inputs have **temporal or sequential structure**
- For tasks involving **context memory**
- Note: Standard RNNs struggle with long-term memory → use LSTM or GRU variants

3. Transformers

Definition

Transformers are deep learning models that handle sequences using **self-attention** mechanisms — no recurrence needed. They enable parallel processing of inputs.

Use Case

- Language models (ChatGPT, BERT, T5)
- Machine translation
- Code generation
- Text summarization

When to Use?

- When working with **language** or **sequential data at scale**
- Need for **long-range dependencies**
- For state-of-the-art performance in **NLP and beyond**