Simple Linear Regression

Summary

• **Simple Linear Regression** is a fundamental machine learning algorithm used for regression problems in supervised learning

• The algorithm creates a **best fit line** to predict continuous output values based on input features

• Simple refers to having **one input feature** (independent variable) and **one output feature** (dependent variable)

• The techniques learned here apply to **deep learning** and **artificial neural networks**

• The best fit line is created by **minimizing the distance** (error) between true data points and predicted points

• Predictions are made by finding where the input value intersects with the best fit line

• Multiple linear regression extends this concept to handle **multiple input features**

Introduction to Simple Linear Regression

Simple Linear Regression is a supervised machine learning algorithm used to solve **regression problem statements**. The techniques learned in this algorithm are foundational and applicable to deep learning, particularly when learning **artificial neural networks**.

The term "simple" refers to the use of **one input feature** (independent variable) to predict **one output feature** (dependent variable). When multiple input features are involved, the technique is called **multiple linear regression**.

Understanding the Problem

Simple linear regression addresses regression problems where we need to predict a continuous output value based on input data.

Example Dataset

Consider a dataset with:

• **Weight** (independent feature)
• **Height** (dependent feature/output)

Sample data points:

• Weight: 74kg → Height: 170cm
• Weight: 80kg → Height: 180cm

• Weight: 75kg → Height: 175.5cm

The goal is to train a model that, given a new weight value, can predict the corresponding height.

The Best Fit Line Concept

Geometric Representation

Simple linear regression creates a **best fit line** through the data points. When plotting weight on the x-axis and height on the y-axis, the algorithm finds a line that best represents the relationship between these variables.

How the Best Fit Line Works

The best fit line is created such that:

• The **distance between true data points** and **predicted points** on the line is minimized

• This distance represents the **error** between actual values and predictions

• The sum of all these distances should be **minimal**

Making Predictions

Once the best fit line is established:

• For a new weight value, draw a vertical line from that point on the x-axis

• Find where this vertical line intersects the best fit line

• The corresponding y-axis value at this intersection point is the **predicted height**

Key Components

Independent Features

The input variable(s) used to make predictions. In simple linear regression, there is **one independent feature**.

Dependent Features

The output variable being predicted. This is the target value that depends on the independent feature(s).

Error Calculation

The error is the distance between:

• **True points**: The actual output values from the dataset (e.g., 170cm, 180cm, 175.5cm)

• **Predicted points**: The values predicted by the best fit line

The algorithm selects the best fit line by minimizing the summation of these distances.

Next Steps

Understanding the **mathematical equation** of the best fit line and the specific **error metric** used will provide deeper insight into how simple linear regression works in practice.