Supervised Learning

Christian Darvin

June 3, 2025

Contents

1 Data

Datasets consist of individual entries that contain **features** (X) and a **label** (y). It similar to a single row in a spreadsheet.
Features: values that a supervised model uses to predict the label.
Label: value that we want to predict.

Feature	**Value**
Operating System	Windows 11
CPU	Intel Core i7-12700H
RAM	16GB DDR4
GPU	NVIDIA RTX 3060
Brand	ASUS
Screen Size and Type	15.6" Full HD IPS
Price ($)	**1,299**

Good datasets are both large (high in quantity) and highly diverse (covering a wide range of categories). Datasets with more features doesn't always produce better predictions because some features might not have significance to the label.

2 Model

A mathematical model that defines the relationship between input feature patterns and output labels.

3 Training

Training a model requires a dataset consisting of input features and their corresponding labels. The objective is to find the best solution for predicting the labels from the features. How do we determine if it's a good solution? By comparing the model's predictions to the actual labels. The difference between them is used to compute the **loss**. This loss guides how the model updates its internal parameters to improve its predictions over time.

Prediction $\rightarrow$ Loss Calculation $\rightarrow$ Parameter Update
Throughout training, we can experiment with different parameters and input features to improve the model's predictions. Feature selection allows us to choose which inputs the model uses.

4 Evaluating

In this stage, we evaluate how well the model performs. We compare the model's prediction to the label's true values.

Prediction $\rightarrow$ Loss Calculation ($\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$)

5 Inference

Once the model has achieved good performance, it can be used to make predictions on unlabeled data.

6 References

- Supervised Learning - Google Developers
- Machine Learning Glossary - Google Developers