

Case Study: Components of Learning

Machine Learning involves a computer program learning from experience (E) with respect to some class of tasks (T) and performance measure (P). This document explores the specific components of learning required to build systems for three distinct use cases.

1. Predicting Placement Chances

Problem Definition

The goal is to build a predictive model that analyzes a student's academic and non-academic profile to estimate the probability of them getting placed in a company.

The Task (T)

Nature: Supervised Learning (Classification/Regression).

Definition: Map input variables (profile) to output (Placement Status).

Output: Binary label (0/1) or probability score.

The Experience / Data (E)

Input Features (X):

- Academic: CGPA, 10th/12th Marks, Arrears.
- Technical: Skills (React, Node.js), Project scores.
- Demographics: Department.

Target Labels (Y): Actual placement status of previous batches.

Performance Measure (P)

- Accuracy: Percentage of correct predictions.
- Precision & Recall: To minimize false positives.
- Confusion Matrix: To visualize classification errors.

Learning Algorithm

- Logistic Regression: For binary classification.
- Random Forest: For handling non-linear data patterns.

2. Detecting Fake Reviews

Problem Definition

The objective is to identify whether a product review posted on an e-commerce platform is genuine or fake (spam/bot-generated).

The Task (T)

Nature: Supervised Learning (NLP).

Definition: Classify text into 'Fake' or 'Genuine'.

The Experience / Data (E)

Input Features (X):

- Textual: Bag-of-words, sentiment score.
- Metadata: Timestamp, account age.
- Behavioral: Rating deviation.

Target Labels (Y): Moderator-flagged reviews.

Performance Measure (P)

- F1-Score: Best for imbalanced data (mostly real reviews).
- ROC-AUC: Measures separation capability.

Learning Algorithm

- Naive Bayes: Probabilistic classifier for text.
- RNNs/LSTMs: Deep learning for context analysis.

3. Predicting Electricity Consumption

Problem Definition

Forecast future electricity consumption based on historical usage and external factors.

The Task (T)

Nature: Regression / Time Series Forecasting.

Definition: Predict a continuous numerical value (kWh).

The Experience / Data (E)

Input Features (X):

- Historical Usage: Previous kWh logs.
- Temporal: Hour, Day, Seasonality.
- External: Temperature, Holidays.

Target Labels (Y): Future consumption values.

Performance Measure (P)

- MAE (Mean Absolute Error): Average prediction error.
- RMSE (Root Mean Squared Error): Penalizes large errors.

Learning Algorithm

- Linear Regression: For simple trends.
- ARIMA: Statistical time-series model.
- LSTM: Neural network for sequential patterns.