Supervised machine learning algorithms

Class 27

6/9/2025

Acknowledgement

The series of the IT & Japanese language course is Supported by AOTS and OEC.

Ministry of Economy, Trade and Industry

Overseas Employment Corporation

What you have Learnt Last Week

We were focused on following points.

- Usage of control and loop flow statement
- Performing Linear Algebra in Numpy
- Software development Life cycle
- Importance of Security compliance, Bash Scripting, Ansible, docker and docker compose
- API testing with Postman and Introduction of Jira
- IAM Permission and S3 bucket
- Introduction to AWS, Azure and GCP

What you will Learn Today

We will focus on following points.

1. Introduction to Supervised Machine Learning
2. Overview of Popular Supervised Machine Learning Algorithms
3. Applications of Supervised Machine learning algorithms
4. Using supervised learning to analyze customer feedback and improve user experience in online platforms
5. Q&A Session

Introduction to Supervised Machine Learning

Definition, Key Concepts & Workflow

Definition: Model learns from labeled data to predict outcomes.

Key Concepts:

•Features (inputs X) → Labels (outputs Y)
•Training Data vs. Testing Data

Comparison:

•Supervised → labeled data
•Unsupervised → unlabeled data
•Reinforcement → rewards & penalties

Workflow: Data Collection → Preprocessing → Training → Evaluation → Deployment

Pros: Accurate, interpretable | **Cons:** Needs lots of labeled data, risk of overfitting

Linear & Logistic Regression

Simple Predictive Models

- **Linear Regression:** Predicts continuous values (e.g., house prices).
- **Logistic Regression:** Predicts binary outcomes (Yes/No, Spam/Not Spam).
- **Strengths:** Simple, easy to interpret.
- **Limitations:** Struggles with complex non-linear data.

Decision Trees, Random Forests & kNN

Tree-based & Instance-based Learning

- **Decision Trees:** Splits data by rules → Easy visualization.
- **Random Forests:** Multiple trees → higher accuracy, less overfitting.
- **k-Nearest Neighbors (kNN):** Classifies based on closest neighbors.
- **Use Cases:** Fraud detection, recommendations, classification tasks.

Support Vector Machines & Gradient Boosting

Advanced Supervised Models

Support Vector Machines (SVM): Finds best boundary (hyperplane) → works well on text/images.

Gradient Boosting (XGBoost, LightGBM, CatBoost): Combines weak learners → very powerful for tabular/structured data.

Strengths: High accuracy, handles complex data.

Limitations: SVM slow on large data, boosting can overfit if not tuned.

Neural Networks for Supervised Learning

Deep Learning Approach

Inspired by human brain → layers of “neurons.”

Handles complex data: images, speech, text.

Examples: Face recognition, speech-to-text, medical imaging.

Strengths: Very powerful on big data.

Limitations: Needs large datasets + high computing power.

Choosing the Right Algorithm

Trade-offs & Best Fit

- **Linear/Logistic Regression:** Simple, fast, explainable.
- **Decision Trees/Random Forests:** Balanced accuracy + interpretability.
- **SVM:** Great for text & image data with clear separation.
- **kNN:** Best for small datasets, simple tasks.
- **Boosting (XGBoost/LightGBM):** Complex structured data, finance, marketing.
- **Neural Networks:** Best for big data (images, speech, NLP).

Applications of Supervised Machine Learning

Real-World Use Cases

- **Healthcare:** Disease prediction, medical imaging.
- **Finance:** Fraud detection, credit scoring, stock trend prediction.
- **E-commerce:** Recommendations, dynamic pricing.
- **Marketing:** Customer segmentation, churn prediction.
- **NLP:** Sentiment analysis, text classification.
- **Manufacturing & IoT:** Predictive maintenance, defect detection.

Using Supervised Learning for Customer Feedback

Collecting & Preprocessing Data

Sources: Reviews, ratings, surveys, chat logs, support tickets.

Preprocessing Steps:

- Remove stopwords, emojis, duplicates.
- Tokenization & lemmatization.
- Convert text into features (TF-IDF, word embeddings).

Ensures clean, usable, and labeled dataset for ML models.

Using Supervised Learning for Customer Feedback

Sentiment Analysis & Opinion Mining

Sentiment Analysis: Classify text as Positive, Negative, or Neutral.

Aspect-based Opinion Mining: Identify opinions on specific features/services.

Example: *“Delivery is slow but packaging is good.”*

Helps companies understand customer pain points.

Using Supervised Learning for Customer Feedback

Churn Prediction & Personalization

Customer Churn Prediction:

- Detect which users are likely to leave.
- Use features: frequency of use, complaints, inactivity.

Retention Strategies: Discounts, loyalty rewards, improved service.

Personalization & Recommendations:

Suggest products/services based on past behavior.
Improves engagement & user satisfaction.

Using Supervised Learning for Customer Feedback

Case Studies & Real-World Use

- **Netflix:** Personalized recommendations using supervised models.
- **Amazon:** Review sentiment analysis for product improvements.
- **Spotify:** Predicts churn and recommends playlists.
- **Airbnb:** Feedback analysis to improve host/guest experience.

Data Preparation for Supervised Learning

Collection, Cleaning & Feature Engineering

Data Collection & Labeling: Labeled examples (positive/negative, churn/no-churn).

Handling Missing Values & Outliers: Replace, drop, or impute values.

Feature Engineering:

- Scaling (Normalization, Standardization).
- Encoding categorical variables (One-hot, Label Encoding).

Good preprocessing → better model accuracy.

Data Preparation for Supervised Learning

Splitting Data & Handling Imbalance

Splitting:

- Train Set (model learns), Validation (tuning), Test Set (final evaluation).
- Use **cross-validation** for robust testing.

Balanced Datasets:

- Prevent bias in imbalanced classes.
- Techniques: Oversampling (SMOTE), Undersampling, Class weights.

- Oversampling = "Make small group bigger“
- Undersampling = "Make big group smaller"

Model Training, Evaluation & Optimization

Metrics, Tuning & Avoiding Overfitting

Training vs. Testing Phase: Train on 70–80%, test on 20–30%.

Evaluation Metrics:

- Classification: Accuracy, Precision, Recall, F1, ROC-AUC.
- Regression: RMSE, MAE.

Hyperparameter Tuning: Grid Search, Random Search, Bayesian Optimization.

Avoiding Overfitting/Underfitting:

Regularization (L1, L2), pruning (trees), dropout (NNs).
Cross-validation for reliability.

Real-World Challenges in Supervised Learning

Data & Interpretability Issues

Data Quality & Labeling:

- Expensive & time-consuming to label data.
- Errors → inaccurate models.

Interpretability vs. Accuracy:

- Simple models (Logistic Regression) = easy to interpret.
- Complex models (Deep Learning) = higher accuracy but “black box.”

Real-World Challenges in Supervised Learning

Scalability, Bias & Deployment

Scalability: Some algorithms (SVM, kNN) struggle with very large datasets.

Ethical Issues & Bias: Models can reflect human biases in data → unfair results.

Deployment Challenges:

- Integrating into real-time production systems.
- Monitoring drift (when new data patterns differ from training data).

Future Trends in Supervised Learning

Advances in Automation & Real-Time Learning

Integration with Deep Learning: Combine ML + neural networks for complex tasks.

AutoML: Automated model selection, hyperparameter tuning, and feature engineering.

Real-Time Applications: Fraud detection, recommendation engines, dynamic pricing.

Future Trends in Supervised Learning

Explainability & Hybrid Approaches

Explainable AI (XAI): Making complex models transparent & interpretable.

Hybrid Approaches:

- Semi-supervised learning (uses small labeled + large unlabeled data).
- Combining reinforcement learning with supervised learning.

- Aim: More accurate, fair, and practical ML models.

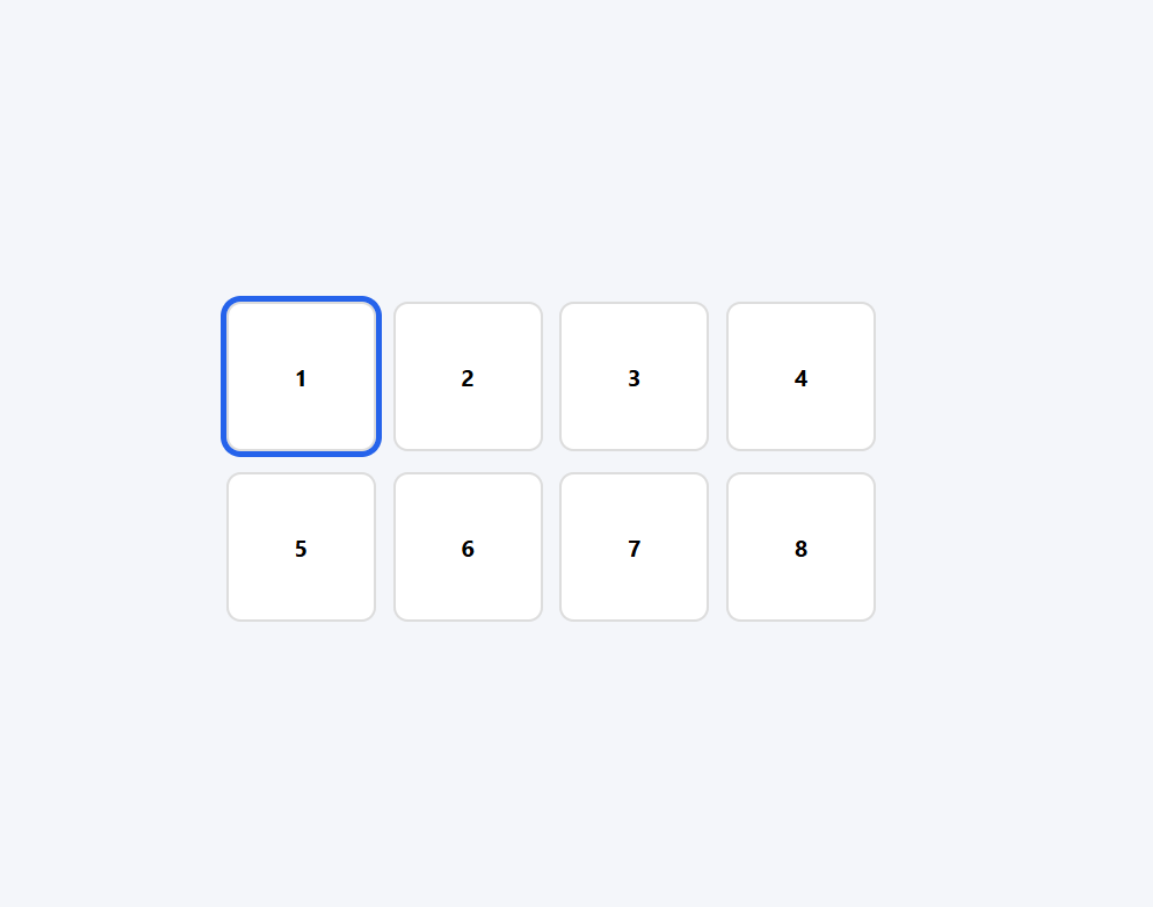
Game 1

Cloud Fundamentals Maze

Step1: Start the Game by Clicking the Link

Step2: Click on the Game It will Start

https://codepen.io/HT-Design/full/WbQPjoL



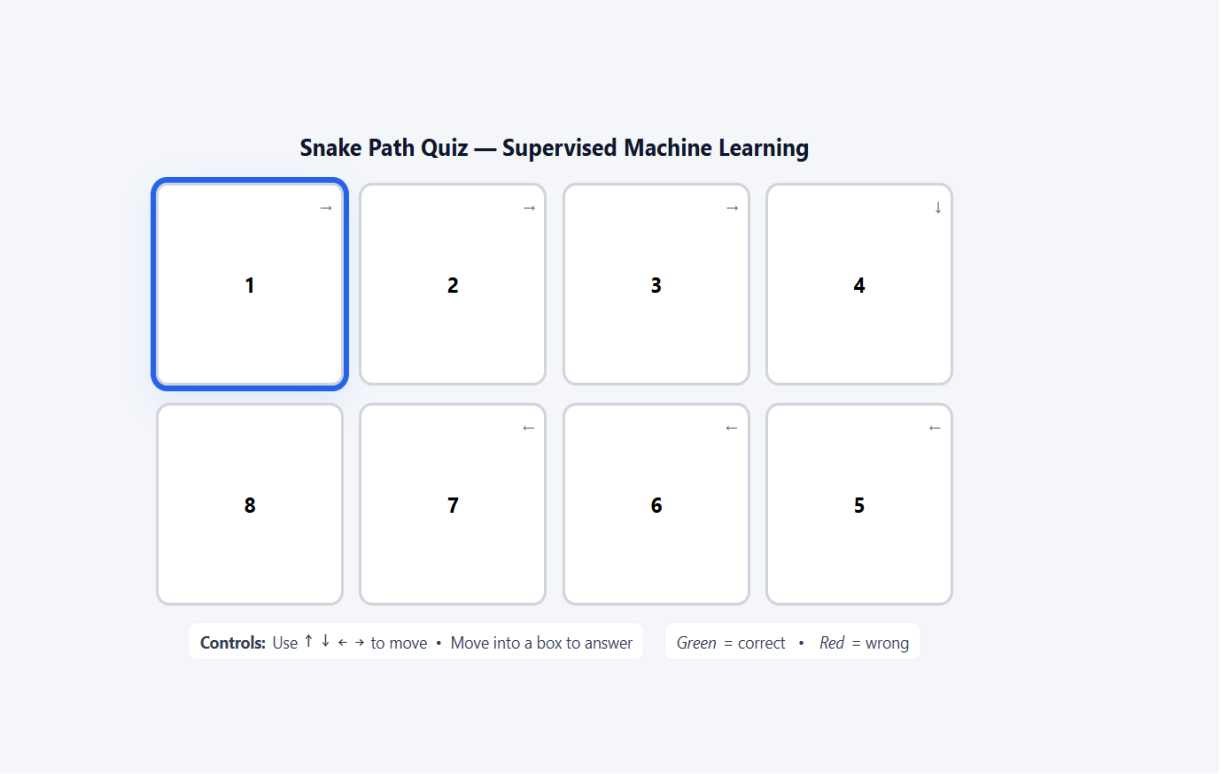
Game 2

Cloud Services Maze

Step1: Start the Game by Clicking the Link

Step2: Click on the Game It will Start

https://codepen.io/HT-Design/full/gbaqWmz



Assignment

Quiz Section

Quiz

Everyone student should click on submit button before time ends otherwise MCQs will not be submitted

[Guidelines of MCQs]

1. There are 20 MCQs
2. Time duration will be 10 minutes
3. This link will be share on 12:25pm (Pakistan time)
4. MCQs will start from 12:30pm (Pakistan time)
5. This is exact time and this will not change
6. Everyone student should click on submit button otherwise MCQs will not be submitted after time will finish
7. Every student should submit Github profile and LinkedIn post link for every class. It include in your performance

Assignment

Assignment should be submit before the next class

[Assignments Requirements]

1. Create a post of today's lecture and post on LinkedIn.
2. Make sure to tag @Plus W @Pak-Japan Centre and instructors LinkedIn profile
3. Upload your code of assignment and lecture on GitHub and share your GitHub profile in respective your region group WhatsApp group
4. If you have any query regarding assignment, please share on your region WhatsApp group.
5. Students who already done assignment, please support other students

Q&A Session

ありがとうございます。
Thank you.
شکریا

For the World with Diverse Individualities