

STA 3180 Statistical Modelling: Machine Learning

STA 3180 Statistical Modelling – Machine Learning Lecture Notes

Introduction

Machine learning is a branch of artificial intelligence that focuses on the development of computer programs that can learn from data and improve their accuracy over time without being explicitly programmed. It is used in a variety of applications, such as image recognition, natural language processing, and predictive analytics.

Key Concepts

- **Supervised Learning:** This type of machine learning involves providing the algorithm with labeled data, which it then uses to make predictions about new data. Examples include regression and classification algorithms.
- **Unsupervised Learning:** This type of machine learning does not require labeled data and instead relies on the algorithm to identify patterns in the data. Examples include clustering and anomaly detection algorithms.
- **Reinforcement Learning:** This type of machine learning involves an agent interacting with its environment and receiving rewards or punishments based on its actions. The agent then uses this feedback to adjust its behavior in order to maximize its rewards.
- **Deep Learning:** This type of machine learning uses neural networks to learn from data. It is often used for tasks such as image recognition and natural language processing.

Definitions

- **Algorithm:** A set of instructions that a computer can follow to solve a problem or accomplish a task.
- **Data:** Information that can be used to make decisions or predictions.
- **Model:** A mathematical representation of a system or process.
- **Training:** The process of using an algorithm to learn from data.
- **Evaluation:** The process of testing a model to determine how well it performs.

Coding Example

Start of Code

```
# Import libraries
```

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

# Load the dataset
data = pd.read_csv('data.csv')

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(data[['x1', 'x2']],
data['y'], test_size=0.2)

# Train the model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
predictions = model.predict(X_test)

End of Code
```

Practice Multiple Choice Questions

Q1. Which of the following is NOT a type of machine learning?

- A. Natural Language Processing
- B. Supervised Learning
- C. Genetic Algorithm
- D. Unsupervised Learning

Answer: C. Genetic Algorithm