Detailed Summary for Machine Learning Exam

# Module 1: Introduction to Machine Learning

## เนื้อหา

1. Concepts of Machine Learning (ML)  
- ML คือการศึกษาเพื่อให้คอมพิวเตอร์เรียนรู้จากข้อมูลโดยไม่ต้องโปรแกรมอย่างชัดเจน  
- Arthur Samuel (1959) ให้คำจำกัดความ ML ว่าเป็นการศึกษาที่ทำให้คอมพิวเตอร์สามารถเรียนรู้ได้เอง  
- ตัวอย่างการใช้งาน ML ได้แก่ Gmail ที่สามารถจำแนกอีเมลเป็นสแปมหรือไม่ และ Facebook ที่สามารถตรวจจับใบหน้าในภาพถ่ายได้  
  
2. Styles of Learning  
- Supervised Learning: การเรียนรู้จากข้อมูลที่มีป้ายกำกับ เช่น การทำนายราคาบ้านจากข้อมูลพื้นที่และจำนวนห้องนอน  
- Unsupervised Learning: การเรียนรู้จากข้อมูลที่ไม่มีป้ายกำกับ เช่น การทำคลัสเตอร์ข้อมูลลูกค้า  
- Reinforcement Learning: การเรียนรู้จากการทำซ้ำและปรับปรุงผลลัพธ์ เช่น การสอนให้คอมพิวเตอร์เล่นเกมโดยใช้การทดลองและความผิดพลาด  
  
3. Basic Terminology and Notations  
- Training example: แถวในตารางที่แสดงชุดข้อมูล ตัวอย่างเช่น การเก็บข้อมูลจากผู้ใช้งาน  
- Feature: คอลัมน์ในตารางข้อมูลที่เรียกว่าตัวทำนายหรืออินพุต เช่น ความสูง น้ำหนัก  
- Target: ค่าผลลัพธ์ที่ต้องการทำนาย เช่น การทำนายว่าเป็นโรคหรือไม่  
- Model: การแทนความสัมพันธ์ระหว่างฟีเจอร์และผลลัพธ์ เช่น y = 2x + 1  
  
4. Machine Learning Systems Workflow  
- การวางแผน: กำหนดขอบเขตและความต้องการของโครงการ  
- การวิเคราะห์: ทำความเข้าใจกับข้อมูลที่มีอยู่  
- การสร้างโมเดล: เลือกวิธีการสร้างโมเดลและปรับปรุงโมเดล  
- การนำไปใช้งาน: นำโมเดลที่สร้างขึ้นไปใช้งานจริง  
  
5. Python for Machine Learning  
- Python เป็นภาษายอดนิยมสำหรับการเรียนรู้ของเครื่องและวิทยาศาสตร์ข้อมูล  
- มีไลบรารีหลายตัวที่ช่วยในการคำนวณทางวิทยาศาสตร์และการสร้างโมเดล เช่น NumPy, Pandas, Matplotlib, scikit-learn, TensorFlow

## วัตถุประสงค์

- เข้าใจแนวคิดพื้นฐานของการเรียนรู้ของเครื่อง  
- เข้าใจและสามารถใช้งาน Python ในการพัฒนาโมเดลการเรียนรู้ของเครื่อง

# Module 2: Data Preprocessing

## เนื้อหา

1. Importance of Data Preprocessing  
- การเตรียมข้อมูลเป็นสิ่งสำคัญเพื่อให้ข้อมูลเหมาะสมกับการใช้ในโมเดลการเรียนรู้ของเครื่อง  
- ช่วยลดความซับซ้อนของข้อมูลและเพิ่มประสิทธิภาพของโมเดล  
  
2. Missing Data Handling  
- วิธีการจัดการข้อมูลที่หายไป เช่น การเติมค่าที่ขาดไป การลบแถวที่มีค่าหายไป  
- การใช้เทคนิคการเติมค่าที่ขาดไป (Imputation) เช่น การเติมค่าเฉลี่ยหรือค่ากลางของข้อมูล  
  
3. Categorical Data Handling  
- การแปลงข้อมูลประเภทที่เป็นข้อความให้เป็นตัวเลข เช่น การใช้ Label Encoding หรือ One-hot Encoding  
- Label Encoding: แทนค่าด้วยตัวเลข เช่น แทน 'ชาย' ด้วย 0 และ 'หญิง' ด้วย 1  
- One-hot Encoding: แยกค่าข้อมูลออกเป็นหลายคอลัมน์ เช่น แทน 'สีแดง' 'สีเขียว' และ 'สีน้ำเงิน' ด้วย 3 คอลัมน์  
  
4. Feature Engineering  
- การปรับปรุงคุณลักษณะของข้อมูลเพื่อเพิ่มประสิทธิภาพการพยากรณ์ของโมเดล  
- การสร้างฟีเจอร์ใหม่จากฟีเจอร์เดิม เช่น การคำนวณอัตราส่วนหรือผลรวมของฟีเจอร์ต่างๆ  
  
5. Data Scaling and Normalization  
- การทำให้ข้อมูลอยู่ในสเกลที่เท่ากัน เช่น การใช้ Normalization หรือ Standardization  
- Normalization: การปรับค่าข้อมูลให้อยู่ในช่วง [0, 1] เช่น การใช้สูตร X\_new = (X - X\_min) / (X\_max - X\_min)  
- Standardization: การปรับค่าข้อมูลให้อยู่ในรูปแบบของค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน เช่น การใช้สูตร X\_new = (X - mean) / std  
  
6. Feature Selection  
- การเลือกคุณลักษณะที่สำคัญเพื่อใช้ในโมเดล เช่น การใช้เทคนิคการเลือกฟีเจอร์ที่มีความสำคัญสูงสุด

## วัตถุประสงค์

- เข้าใจและสามารถเลือกวิธีการจัดการกับข้อมูลที่ขาดหายและข้อมูลประเภท  
- สามารถใช้เทคนิคการวิศวกรรมคุณลักษณะเพื่อปรับปรุงข้อมูลสำหรับกรณีศึกษา

# Module 3: Supervised Learning

## เนื้อหา

1. Supervised Learning Concepts  
- การเรียนรู้จากข้อมูลที่มีป้ายกำกับ เช่น การทำนายราคาบ้านจากข้อมูลพื้นที่และจำนวนห้องนอน  
- การใช้ข้อมูลที่มีคำตอบถูกต้องเพื่อสอนโมเดล  
  
2. Supervised Learning Tasks  
- Regression: การทำนายค่าต่อเนื่อง เช่น ราคา  
- Classification: การทำนายค่าที่เป็นประเภท เช่น การจำแนกโรค  
  
3. Supervised Learning Algorithms  
- Regression: Linear Regression, Nonlinear Regression  
- Linear Regression: การทำนายค่าด้วยเส้นตรง เช่น y = 2x + 1  
- Nonlinear Regression: การทำนายค่าด้วยเส้นโค้ง เช่น การใช้ Polynomial Regression  
- Classification: Logistic Regression, Naive Bayes, Decision Tree, Gradient Boosting, K-nearest neighbors, Support Vector Machine, Artificial Neural Networks  
- Logistic Regression: การทำนายค่าที่เป็นประเภทด้วยโมเดลเชิงเส้น  
- Decision Tree: การใช้โครงสร้างต้นไม้ในการทำนายค่า  
  
4. Regression Model Evaluation  
- การประเมินโมเดลเชิงเส้น เช่น R-squared, Mean Squared Error  
- R-squared: ค่าที่บ่งบอกถึงความสัมพันธ์ระหว่างฟีเจอร์และผลลัพธ์  
- Mean Squared Error (MSE): ค่าที่บ่งบอกถึงความแตกต่างระหว่างค่าที่ทำนายและค่าจริง

## วัตถุประสงค์

- เข้าใจและสามารถเลือกใช้อัลกอริธึมการเรียนรู้ที่ควบคุมได้ให้เหมาะสมกับกรณีศึกษา  
- สามารถประเมินประสิทธิภาพของโมเดลการทำนายได้

# Detailed Explanation for Supervised Learning Algorithms

## Linear Regression

การถดถอยเชิงเส้น (Linear Regression) เป็นวิธีการเรียนรู้ที่ใช้ในการทำนายค่าต่อเนื่อง (Continuous Value) โดยมีสมมติฐานว่าความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) และตัวแปรตาม (Dependent Variable) เป็นเส้นตรง สูตรพื้นฐานของ Linear Regression คือ:  
y = β0 + β1 x  
โดยที่:  
- y คือค่าที่ต้องการทำนาย  
- x คือค่าของตัวแปรอิสระ  
- β0 คือค่าคงที่ (Intercept)  
- β1 คือสัมประสิทธิ์ของตัวแปรอิสระ (Slope)

## Logistic Regression

การถดถอยโลจิสติก (Logistic Regression) ใช้สำหรับการทำนายค่าที่เป็นประเภท (Categorical Value) โดยเฉพาะการทำนายค่าทางสองประเภท (Binary Classification) เช่น การทำนายว่าผู้ป่วยมีโรคหรือไม่ (มีโรค/ไม่มีโรค) โดยใช้ฟังก์ชันโลจิสติก (Logistic Function) หรือฟังก์ชัน Sigmoid ซึ่งมีสมการคือ:  
P(y=1|x) = 1 / (1 + e^-(β0 + β1 x))  
โดยที่:  
- P(y=1|x) คือความน่าจะเป็นที่ y เป็น 1 เมื่อมีค่า x  
- β0 และ β1 คือพารามิเตอร์ของโมเดล

## Decision Tree

ต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีการเรียนรู้ที่ใช้โครงสร้างต้นไม้ในการทำนายค่า โดยการแบ่งข้อมูลออกเป็นกลุ่มย่อย ๆ ตามคุณลักษณะของข้อมูล เพื่อให้ได้ผลลัพธ์ที่เป็นไปตามข้อกำหนด โดยมีการเลือกคุณลักษณะที่จะใช้แบ่งข้อมูลโดยพิจารณาจากการลดความไม่แน่นอน (Uncertainty) หรือการเพิ่มความบริสุทธิ์ (Purity) ของกลุ่มย่อย ๆ ที่ได้จากการแบ่ง

## Gradient Boosting

การเพิ่มความเข้มข้นแบบกราดิเอนต์ (Gradient Boosting) เป็นวิธีการเรียนรู้ที่ใช้การสร้างโมเดลหลาย ๆ โมเดลที่มีความอ่อนแอ (Weak Learners) มารวมกันเพื่อสร้างโมเดลที่มีความเข้มแข็ง (Strong Learner) โดยการสร้างโมเดลแต่ละตัวจะมีการปรับปรุงข้อผิดพลาดของโมเดลก่อนหน้า

## Support Vector Machine (SVM)

เครื่องเวกเตอร์สนับสนุน (Support Vector Machine - SVM) เป็นวิธีการเรียนรู้ที่ใช้สำหรับการจำแนกประเภท โดยการหาขอบเขต (Hyperplane) ที่สามารถแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้อย่างชัดเจน โดยการเลือกขอบเขตที่มีระยะห่างมากที่สุดระหว่างกลุ่มข้อมูลที่อยู่ใกล้ขอบเขตมากที่สุด (Support Vectors)

## Artificial Neural Networks (ANN)

โครงข่ายประสาทเทียม (Artificial Neural Networks - ANN) เป็นวิธีการเรียนรู้ที่เลียนแบบการทำงานของสมองมนุษย์ โดยใช้โหนด (Nodes) หรือเซลล์ประสาทเทียม (Artificial Neurons) ที่เชื่อมต่อกันเป็นชั้น ๆ (Layers) และมีการปรับน้ำหนัก (Weights) ของการเชื่อมต่อเพื่อให้ได้ผลลัพธ์ที่ต้องการ