機器學習第一章：考試筆記與模擬出題

# 📘 第一章：機器學習導論（Introduction to Machine Learning）

## 🧠 名詞解釋（Term Explanation）

### 機器學習（Machine Learning）

中文：機器學習是一種人工智慧（AI）的子領域，旨在讓電腦能從資料中自動學習模式，而非直接透過程式碼規則進行推論。它能從樣本中建立模型，用來預測或分類新資料。

英文：Machine Learning is a subfield of AI that enables computers to learn patterns from data without being explicitly programmed. It builds models from sample data to make predictions or classifications on new data.

### 監督式學習（Supervised Learning）

中文：一種基於「有標記資料」進行訓練的方法。模型學習輸入與正確輸出間的映射，用來預測新資料的輸出。常見應用包括分類與回歸。

英文：Supervised learning trains a model using labeled data where each input is paired with the correct output. The goal is to learn a mapping to predict the output of new data. Common tasks include classification and regression.

### 非監督式學習（Unsupervised Learning）

中文：使用未標記資料尋找隱含結構或模式，例如分群、降維。常用於探索性資料分析。

英文：Unsupervised learning works with unlabeled data to find hidden patterns or structures, such as clustering or dimensionality reduction.

### 半監督式學習（Semi-Supervised Learning）

中文：結合少量標記資料與大量未標記資料進行學習。常見於資料標記困難或昂貴的場景中，如圖像辨識。

英文：Semi-Supervised Learning combines a small amount of labeled data with a large amount of unlabeled data. It is useful when labeling data is expensive or difficult.

### 強化學習（Reinforcement Learning, RL）

中文：代理人透過與環境互動學習最佳行動策略，以最大化長期報酬。核心元素包含狀態（State）、行動（Action）、回饋（Reward）。

英文：Reinforcement Learning is a learning paradigm where an agent interacts with an environment to learn an optimal policy that maximizes long-term rewards.

### 感知器（Perceptron）

中文：一種最早期的人工神經元模型，由輸入的加權總和經過激活函數產生輸出，是現代神經網絡的基礎。

英文：A perceptron is one of the earliest artificial neuron models. It computes a weighted sum of inputs followed by an activation function.

### 回歸（Regression）

中文：一種監督式學習方法，用來預測連續型目標變數。常用於價格預測、趨勢分析等。

英文：Regression is a type of supervised learning method used to predict continuous target variables.

### 遷移學習（Transfer Learning）

中文：將已訓練模型的知識應用到不同但相關的新任務上，可減少資料與訓練時間需求。

英文：Transfer learning applies knowledge from a pre-trained model to a different but related task.

### 生成對抗網路（GANs）

中文：由生成器與判別器構成的深度學習模型，用於生成與真實資料分佈相似的新樣本。

英文：GANs are deep learning models consisting of a generator and a discriminator that produce new samples similar to real data.

### 多任務學習（Multi-Task Learning, MTL）

中文：一個模型同時學習多個相關任務，透過共享表示提升學習效能。

英文：Multi-Task Learning enables a single model to learn multiple related tasks simultaneously by sharing representations.

### 可解釋性與LIME（Explainability & LIME）

中文：可解釋性指模型預測的可理解程度。LIME 是一種用於解釋任意模型預測結果的技術。

英文：Explainability refers to how understandable a model's predictions are. LIME is a technique for explaining predictions of any model.

## 📚 名詞比較（Term Comparison）

Supervised Learning vs. Unsupervised Learning

|  |  |  |
| --- | --- | --- |
| 項目 | Supervised Learning（監督式學習） | Unsupervised Learning（非監督式學習） |
| 資料類型 | 有標記資料（labeled data） | 無標記資料（unlabeled data） |
| 任務目標 | 預測明確輸出，如分類或回歸 | 發現資料中潛在的結構 |
| 常見方法 | 線性回歸、邏輯回歸、SVM、神經網絡 | K-means、PCA、自編碼器 |
| 應用範例 | 電子郵件分類、房價預測 | 客戶分群、新聞分類、異常偵測 |
| 難度 | 標記成本高但訓練明確 | 資料便宜但模型難評估 |

# 📝 第一章模擬出題（Chapter 1 Mock Questions）

## 考題 1

中文問題：什麼是機器學習（Machine Learning）？它與傳統程式設計有何不同？

英文問題：What is Machine Learning (ML)? How does it differ from traditional programming?

中文答案：機器學習是一種人工智慧方法，透過從資料中學習模型來預測結果，不需要人工明確寫出規則。相較之下，傳統程式設計依賴開發者明確定義每個邏輯條件。

英文答案：Machine Learning is an AI approach where models learn patterns from data to make predictions, without being explicitly programmed with rules.

## 考題 2

中文問題：比較監督式學習與非監督式學習的異同。

英文問題：Compare and contrast Supervised and Unsupervised Learning.

中文答案：監督式學習使用有標記資料，訓練模型進行分類或回歸；非監督式學習使用未標記資料，找出資料的內在結構如分群。前者易於評估，但標記成本高；後者資料便宜，但難以驗證結果。

英文答案：Supervised learning uses labeled data to train models for classification or regression. Unsupervised learning finds hidden structures from unlabeled data, such as clustering.

## 考題 3

中文問題：什麼是強化學習（Reinforcement Learning）？其核心概念有哪些？

英文問題：What is Reinforcement Learning (RL)? What are its core components?

中文答案：強化學習是一種透過「試誤」方式從環境中學習最佳決策策略的方法。其核心包含：狀態（state）、行動（action）、回饋（reward）、政策（policy）。

英文答案：Reinforcement Learning is a trial-and-error based learning method where an agent learns optimal decision-making by interacting with an environment.

## 考題 4

中文問題：舉例說明自駕車如何同時應用監督式、非監督式與深度學習。

英文問題：Give an example of how self-driving cars use supervised, unsupervised, and deep learning together.

中文答案：監督式學習用於影像辨識如紅綠燈偵測；非監督式學習用於分析駕駛行為模式；深度學習則用於複雜視覺任務如車道線辨識與物件追蹤。

英文答案：Supervised learning is used for image recognition such as traffic light detection. Unsupervised learning helps analyze driving behavior patterns.

## 考題 5

中文問題：解釋什麼是「遷移學習」（Transfer Learning）及其優點。

英文問題：Explain what Transfer Learning is and mention its advantages.

中文答案：遷移學習是在新任務中利用在舊任務上已訓練好的模型知識，降低資料需求並加速訓練。適用於資料有限或類似任務的場景。

英文答案：Transfer Learning uses knowledge from a pre-trained model on a previous task to improve learning on a new task.

## 考題 6

中文問題：什麼是 GANs（生成對抗網路）？其運作原理為何？

英文問題：What are GANs (Generative Adversarial Networks)? How do they work?

中文答案：GAN 是由生成器與鑑別器兩個模型組成的架構，兩者互相對抗以產生逼真的資料。生成器試圖騙過鑑別器，鑑別器則努力辨識真偽。

英文答案：GANs consist of a generator and a discriminator in a competitive setting. The generator tries to produce realistic data to fool the discriminator.

## 考題 7

中文問題：比較 Transfer Learning 與 Multi-task Learning 的差異。

英文問題：Compare Transfer Learning and Multi-task Learning.

中文答案：Transfer Learning 是將舊任務知識轉移到新任務上，先訓練再微調；Multi-task Learning 是同時訓練多任務，共享特徵提升效率。

英文答案：Transfer Learning transfers knowledge from one task to another. Multi-task Learning trains multiple tasks simultaneously using shared features.

## 考題 8

中文問題：說明為何在深度模型中需要「可解釋性」，並舉出一種方法。

英文問題：Why is explainability important in deep models? Name one explainability method.

中文答案：深度模型如黑箱，不易解釋預測依據，可能導致偏誤或信任危機。LIME 是常見方法，可針對特定預測提供人類可解釋的區域性線性模型。

英文答案：Deep models are black-boxes and difficult to interpret. LIME explains individual predictions using interpretable local linear models.

## 考題 9

中文問題：何謂「回歸」（Regression）？與分類有何不同？

英文問題：What is Regression? How does it differ from Classification?

中文答案：回歸是一種預測連續數值（如價格）的監督學習方法；分類則是預測離散類別（如貓或狗）。

英文答案：Regression predicts continuous values (e.g., prices), while classification predicts discrete categories (e.g., cat or dog).

## 考題 10

中文問題：近年來有哪些機器學習的重大進展？請舉出三種並簡要說明。

英文問題：What are some recent advances in Machine Learning? Name three and briefly explain.

中文答案：1. 遷移學習：節省資料並加速訓練。  
2. GANs：用於生成影像、音訊與資料增強。  
3. 解釋性技術（如LIME）：提高模型可理解性與透明性。

英文答案：1. Transfer Learning: reduces data/training time.  
2. GANs: for image/audio generation and data augmentation.  
3. Explainability (e.g., LIME): enhances transparency of model decisions.

## 📝 額外模擬出題（Chapter 1 Mock Questions – Updated Format）

1. Q1. (a)(b) 題組：機器學習與傳統程式設計的比較

* (a) Explain the main difference between Machine Learning and traditional programming with a concrete example.  
  (b) In the context of email spam filtering, explain how supervised and unsupervised learning would approach the task differently.
* （a）請解釋機器學習與傳統程式設計的主要差異，並舉一實例說明。  
  （b）若應用在電子郵件垃圾信分類上，監督式與非監督式學習會如何採取不同方法？

Q1 標準答案

* English:  
  (a) Machine Learning allows the model to learn patterns from data to make predictions, unlike traditional programming which requires explicitly defined rules. For example, a spam filter using ML learns from labeled emails, while a rule-based program uses hand-coded filters.  
  (b) Supervised learning uses labeled spam/not-spam emails to build a classifier. Unsupervised learning might cluster emails based on word usage and uncover spam-like groups without labels.
* 中文：  
  （a）機器學習能從資料中學習模式進行預測，而非傳統程式設計須由人手動編寫規則。例如垃圾信過濾器，機器學習從標記過的郵件中學習，而傳統程式會硬編規則。  
  （b）監督式學習使用已標記為垃圾信與非垃圾信的資料來訓練分類器；非監督式學習則可從字詞頻率等資料自動將信件分群找出可疑類別。

1. Q2. (a)(b) 題組：不同學習型態的應用

* (a) Compare Supervised, Unsupervised, and Reinforcement Learning in terms of data usage and application scenario.  
  (b) For each type, give one real-world example in the context of self-driving cars.
* （a）請比較監督式學習、非監督式學習與強化學習在資料使用與應用場景上的不同。  
  （b）請分別舉出這三種學習在自駕車中的應用實例。

Q2 標準答案

* English:  
  (a) Supervised learning: labeled data, e.g., image classification. Unsupervised learning: unlabeled data, e.g., customer segmentation. Reinforcement learning: interaction-based, e.g., a robot learning to walk.  
  (b) In self-driving cars: Supervised - object detection; Unsupervised - behavior clustering; Reinforcement - decision making for path planning.
* 中文：  
  （a）監督式學習使用有標記資料（如影像分類）；非監督式學習使用無標記資料（如顧客分群）；強化學習基於回饋互動（如機器人學走路）。  
  （b）自駕車中：監督式用於偵測交通號誌與物件；非監督式可分析駕駛行為模式；強化學習則負責路徑決策與控制。

1. Q3. (a)(b) 題組：進階學習技術與差異分析

* (a) What is Transfer Learning? Why is it useful when labeled data is limited?  
  (b) How does Transfer Learning differ from Multi-Task Learning in the model training process?
* （a）什麼是遷移學習？當標記資料稀少時，為何它特別有用？  
  （b）遷移學習與多任務學習在模型訓練方式上有何不同？

Q3 標準答案

* English:  
  (a) Transfer learning reuses a model trained on one task for a new but related task, useful when labeled data is scarce.  
  (b) Transfer learning trains sequentially (source then target), while Multi-task learning trains multiple tasks simultaneously with shared layers.
* 中文：  
  （a）遷移學習將原任務訓練好的模型知識應用到新任務中，當標記資料稀少時尤其有效。  
  （b）遷移學習是先訓練再微調；多任務學習則是同時訓練多個任務，並共享模型特徵。

1. Q4. (a)(b) 題組：模型可解釋性

* (a) Why is explainability important in Machine Learning? Give one risk of having a black-box model.  
  (b) Briefly describe how LIME helps make model predictions more understandable.
* （a）為何機器學習模型的可解釋性很重要？請說明黑箱模型可能帶來的風險。  
  （b）簡要說明 LIME 如何幫助解釋模型的預測結果。

Q4 標準答案

* English:  
  (a) Explainability is key to trust in ML models. A black-box model may make biased decisions that users cannot audit.  
  (b) LIME creates a simple interpretable model locally around the prediction to explain which features influenced the result.
* 中文：  
  （a）可解釋性可提升對模型的信任。黑箱模型可能有偏誤決策但無法被理解與監督。  
  （b）LIME 可在模型附近擬合一個簡單的可解釋模型來說明特徵對預測的影響程度。

1. Q5. (a)(b) 題組：監督式學習與分類任務的圖像理解

* (a) Based on a labeled dataset of home sizes and prices, explain how a supervised learning algorithm can be trained to predict house prices. Refer to the idea shown in the regression figure in the lecture.  
  (b) Suppose the dataset includes outliers. How might those affect the prediction model, and what preprocessing steps could mitigate their impact?
* （a）根據一組帶有標記的房屋大小與價格資料，說明監督式學習如何建立房價預測模型，並參照課堂上提到的回歸圖示意。  
  （b）若資料中存在異常值，這些資料將如何影響模型？可採取哪些預處理措施減緩其影響？

Q5 標準答案

* English:  
  (a) A supervised regression model fits a line through training data (house size vs. price) to predict price for new sizes. The model minimizes the prediction error.  
  (b) Outliers (e.g., unusually cheap or expensive homes) skew the line. We can reduce their impact via outlier detection or using robust regression.
* 中文：  
  （a）監督式回歸模型可根據訓練資料（房屋大小與價格）擬合預測線，來預測新房屋價格，目標是最小化預測誤差。  
  （b）異常值會拉偏整體趨勢，建議透過異常值偵測或使用健壯回歸法來減弱其影響。

## 📊 模擬題對應圖片

圖 Q5：Supervised Learning - 房價預測回歸圖

