機器學習第五章：完整筆記與模擬出題（未壓縮）

本文件包含梯度下降與交叉驗證的名詞解釋、名詞比較與模擬出題（中英文對照），所有內容逐字保留未壓縮

# 📘 第五章：梯度下降與交叉驗證（Gradient Descent and Cross-Validation）

## 🧠 名詞解釋（Term Explanation）

### 梯度下降（Gradient Descent, GD）

中文：梯度下降是一種優化方法，透過計算損失函數的梯度，並往其反方向移動以最小化誤差。

英文：Gradient descent is an optimization method that minimizes error by computing the gradient of the loss function and moving in the opposite direction.

### 學習率（Learning Rate, η）

中文：學習率決定每次參數更新的步長。太小會導致收斂慢，太大可能震盪或發散。

英文：The learning rate determines the step size of each update. Too small leads to slow convergence; too large may cause oscillation or divergence.

### 損失函數（Loss Function / Cost Function）

中文：衡量模型預測值與真實值差異的函數，是訓練時要最小化的對象。

英文：A function that measures the discrepancy between predicted values and actual targets; the goal of training is to minimize it.

### 隨機梯度下降（Stochastic Gradient Descent, SGD）

中文：是一種梯度下降方法，每次只使用一筆資料來更新參數，更新速度快但路徑不穩定。

英文：A variant of gradient descent that updates the parameters using one data point at a time. It converges faster but with a noisier path.

### 小批次梯度下降（Mini-Batch Gradient Descent）

中文：是 GD 與 SGD 的折衷，每次更新使用一小批資料（如32筆），能結合穩定性與速度。

英文：A compromise between GD and SGD, updating parameters using small batches (e.g., 32 samples), combining stability and speed.

### 批次（Batch）與紀元（Epoch）

中文：一個批次是一次使用的資料子集；一個紀元是整筆訓練資料被使用過一次的週期。

英文：A batch is a subset of data used in one update; an epoch means one full pass through the entire training dataset.

### 早停（Early Stopping）

中文：當驗證集誤差不再下降時，提前停止訓練以避免過擬合。

英文：Stops training when validation error stops decreasing to prevent overfitting.

### 交叉驗證（Cross-Validation）

中文：將資料分成多個子集，多次訓練與驗證模型以獲得穩定的評估結果。

英文：Splits data into multiple folds, training and validating across folds to obtain a more stable performance estimate.

### K折交叉驗證（K-Fold Cross-Validation）

中文：將資料平均分成 K 個子集，每次用 K-1 個做訓練，剩下的做驗證，重複 K 次。

英文：Divides data into K equal subsets; each subset is used once as validation while others train the model.

### 留一交叉驗證（Leave-One-Out Cross Validation, LOOCV）

中文：K-Fold 的特例，K 等於資料總數。每次只留一筆資料做驗證。

英文：A special case of K-Fold where K equals the total number of data points. Each data point is used once for validation.

### 分層交叉驗證（Stratified Cross-Validation）

中文：適用於分類任務，確保每一折中的類別比例與原始資料集相同。

英文：Ensures that the class proportions in each fold are representative of the whole dataset, ideal for classification problems.

### 時序交叉驗證（Time-Series Cross-Validation）

中文：適用於時間序列資料，保留資料的時間順序，不允許未來資料用來預測過去。

英文：Used for time-series data where temporal order must be preserved; prevents leakage from future to past.

## 📚 名詞比較（Term Comparison）

GD vs. SGD vs. Mini-Batch GD

|  |  |  |  |
| --- | --- | --- | --- |
| 項目 | Gradient Descent (GD) | Stochastic Gradient Descent (SGD) | Mini-Batch Gradient Descent |
| 資料使用量 | 每次使用全部資料 | 每次使用 1 筆資料 | 每次使用一小批資料 |
| 收斂速度 | 慢 | 快，但不穩 | 快且較穩 |
| 計算效率 | 高成本 | 低成本 | 平衡 |
| 適用場景 | 資料量小 | 資料量大 | 深度學習常用 |

K-Fold vs. Stratified K-Fold vs. LOOCV

|  |  |  |  |
| --- | --- | --- | --- |
| 項目 | K-Fold | Stratified K-Fold | LOOCV |
| 分割方式 | 隨機等分 | 保留類別比例 | 每次留一筆資料 |
| 適用任務 | 一般任務 | 分類任務 | 小樣本學習 |
| 優點 | 平均化偏誤與變異數 | 分類結果穩定 | 最大化利用資料，偏誤小 |
| 缺點 | 不適合不平衡資料 | 計算較慢 | 計算成本極高 |

# 📝 第五章模擬出題（Chapter 5 Mock Questions）

## 考題 1

中文問題：什麼是梯度下降？它的運作原理是什麼？

英文問題：What is gradient descent and how does it work?

中文答案：梯度下降是一種優化方法，透過計算損失函數的梯度，並往其反方向移動以最小化誤差。

英文答案：Gradient descent is an optimization method that minimizes error by computing the gradient of the loss function and moving in the opposite direction.

## 考題 2

中文問題：什麼是學習率？如果學習率設得太大或太小，會發生什麼問題？

英文問題：What is the learning rate? What happens if it is too large or too small?

中文答案：學習率決定每次參數更新的步長。太小會導致收斂慢，太大可能震盪或發散。

英文答案：The learning rate determines the step size of each update. Too small leads to slow convergence; too large may cause oscillation or divergence.

## 考題 3

中文問題：請比較 GD、SGD 與 Mini-Batch 的異同。

英文問題：Compare GD, SGD, and Mini-Batch Gradient Descent.

中文答案：GD 使用所有資料，SGD 每次用一筆資料，Mini-Batch 用小批資料更新。Mini-Batch 是效率與穩定性的折衷。

英文答案：GD uses all data, SGD uses one sample, and Mini-Batch uses a small batch. Mini-Batch is a trade-off between speed and stability.

## 考題 4

中文問題：如果你使用 SGD 觀察到訓練過程震盪劇烈，你會採取哪些措施？

英文問題：If you observe severe oscillation during SGD training, what could you do?

中文答案：可嘗試降低學習率、改用 Mini-Batch 或加入動量項（Momentum）來平滑路徑。

英文答案：You may reduce the learning rate, switch to Mini-Batch, or add momentum to smooth the updates.

## 考題 5

中文問題：什麼是交叉驗證？為什麼在模型訓練中重要？

英文問題：What is cross-validation and why is it important?

中文答案：交叉驗證將資料分折評估模型效能，避免訓練資料過擬合。能提升泛化能力的估計準確度。

英文答案：Cross-validation splits data into folds to evaluate model performance and avoid overfitting. It improves estimation of generalization ability.

## 考題 6

中文問題：請比較 K-Fold 與 Stratified K-Fold。

英文問題：Compare K-Fold and Stratified K-Fold cross-validation.

中文答案：K-Fold 隨機分組可能導致類別比例不一致；Stratified K-Fold 保證每一折類別比例相同，適合分類任務。

英文答案：K-Fold randomly splits data, which may cause class imbalance; Stratified K-Fold ensures equal class proportions in each fold, ideal for classification.

## 考題 7

中文問題：在分類問題中，如果資料分布不平衡，你應該使用哪種交叉驗證方法？

英文問題：In classification with imbalanced data, which cross-validation method should you use?

中文答案：應使用 Stratified K-Fold，以確保各折類別比例一致，避免評估偏誤。

英文答案：Use Stratified K-Fold to maintain class balance across folds and avoid biased evaluation.

## 考題 8

中文問題：什麼是早停（Early Stopping）？有什麼優點？

英文問題：What is early stopping and what is its advantage?

中文答案：早停是在驗證集誤差開始上升前停止訓練，可避免過擬合並節省計算資源。

英文答案：Early stopping halts training before validation error rises, preventing overfitting and saving computation.

## 考題 9

中文問題：留一交叉驗證（LOOCV）有什麼優缺點？

英文問題：What are the pros and cons of Leave-One-Out Cross-Validation (LOOCV)?

中文答案：優點是偏差小，資料充分利用；缺點是需訓練 n 次模型，計算成本高。

英文答案：It has low bias and fully utilizes data, but it requires training n models, making it computationally expensive.

## 考題 10

中文問題：為什麼時間序列資料不適用一般 K-Fold 驗證？應該使用什麼方法？

英文問題：Why is regular K-Fold not suitable for time series? What method should you use?

中文答案：時間序列有時間先後順序，不能用未來資料預測過去。應使用 Time-Series Cross-Validation。

英文答案：Time series has temporal order, so future data cannot predict past. Time-Series Cross-Validation should be used.

# 📝 第五章：模擬考題與答案（題目與答案寫在一起）

1. Q1. (a)(b) 題組：梯度下降與學習率

(a) What is gradient descent and how does it work?  
Answer: Gradient descent is an optimization method that minimizes error by computing the gradient of the loss function and moving in the opposite direction.  
（a）什麼是梯度下降？它的運作原理是什麼？  
答：梯度下降是一種優化方法，透過計算損失函數的梯度，並往其反方向移動以最小化誤差。  
  
(b) What is the learning rate? What happens if it is too large or too small?  
Answer: The learning rate determines the step size of each update. Too small leads to slow convergence; too large may cause oscillation or divergence.  
（b）什麼是學習率？如果設得太大或太小，會發生什麼問題？  
答：學習率決定每次參數更新的步長。太小會導致收斂慢，太大可能震盪或發散。

1. Q2. (a)(b) 題組：GD, SGD, Mini-Batch 比較與應用

(a) Compare GD, SGD, and Mini-Batch Gradient Descent.  
Answer: GD uses all data, SGD uses one sample, and Mini-Batch uses a small batch. Mini-Batch is a trade-off between speed and stability.  
（a）請比較 GD、SGD 與 Mini-Batch 的異同。  
答：GD 使用所有資料，SGD 每次用一筆資料，Mini-Batch 用小批資料更新。Mini-Batch 是效率與穩定性的折衷。  
  
(b) If you observe severe oscillation during SGD training, what could you do?  
Answer: You may reduce the learning rate, switch to Mini-Batch, or add momentum to smooth the updates.  
（b）如果你使用 SGD 觀察到訓練過程震盪劇烈，你會採取哪些措施？  
答：可嘗試降低學習率、改用 Mini-Batch 或加入動量項（Momentum）來平滑路徑。

1. Q3. (a)(b) 題組：交叉驗證概念與方法

(a) What is cross-validation and why is it important?  
Answer: Cross-validation splits data into folds to evaluate model performance and avoid overfitting. It improves estimation of generalization ability.  
（a）什麼是交叉驗證？為什麼在模型訓練中重要？  
答：交叉驗證將資料分折評估模型效能，避免訓練資料過擬合。能提升泛化能力的估計準確度。  
  
(b) Compare K-Fold and Stratified K-Fold cross-validation.  
Answer: K-Fold randomly splits data, which may cause class imbalance; Stratified K-Fold ensures equal class proportions in each fold, ideal for classification.  
（b）請比較 K-Fold 與 Stratified K-Fold。  
答：K-Fold 隨機分組可能導致類別比例不一致；Stratified K-Fold 保證每一折類別比例相同，適合分類任務。

1. Q4. (a)(b) 題組：特殊交叉驗證方法與應用

(a) In classification with imbalanced data, which cross-validation method should you use?  
Answer: Use Stratified K-Fold to maintain class balance across folds and avoid biased evaluation.  
（a）在分類問題中，如果資料分布不平衡，你應該使用哪種交叉驗證方法？  
答：應使用 Stratified K-Fold，以確保各折類別比例一致，避免評估偏誤。  
  
(b) Why is regular K-Fold not suitable for time series? What method should you use?  
Answer: Time series has temporal order, so future data cannot predict past. Time-Series Cross-Validation should be used.  
（b）為什麼時間序列資料不適用一般 K-Fold 驗證？應該使用什麼方法？  
答：時間序列有時間先後順序，不能用未來資料預測過去。應使用 Time-Series Cross-Validation。

1. Q5. (a)(b) 題組：學習率與收斂路徑（圖：learning\_rate\_effect.png）

(a) According to the figure, how does the learning rate affect the convergence behavior of gradient descent?  
Answer: A too small learning rate causes slow convergence. A good learning rate leads to stable descent. A too large learning rate causes the loss to oscillate and may even diverge.  
（a）根據圖中學習率變化對收斂影響的結果，請說明其意義。  
答：學習率太小會導致收斂速度非常慢；適當學習率能穩定收斂；學習率過大會導致震盪甚至發散。  
  
(b) In practical training, how can you determine whether the learning rate needs adjustment?  
Answer: If the loss decreases very slowly, increase the learning rate slightly. If the loss oscillates or increases, reduce the learning rate.  
（b）在實際訓練中，如何判斷是否需要調整學習率？  
答：若損失下降非常慢，可略微增加學習率；若損失震盪或上升，應減少學習率。

圖 Q5：Effect of Learning Rate on Convergence

