

iPAS AI 應用規劃師（初級）

科目一：人工智慧基礎概論

考前衝刺版

把原本完整講義濃縮成「考前反覆看、快速背、快速判斷」版本

本版依據：前一版由使用者提供 iPAS 學習指引整理出的講義。原文件定位為學習指引，不是正式教材或題庫，因此以下內容以原講義已整理出的考試範圍與術語為基礎，不宜稱為實際考題。



一、考試範圍先背這 4 大塊

① AI 基礎概念	② 資料處理與分析	③ 機器學習	④ 鑑別式／生成式 AI
AI 定義、分類、應用、治理	資料類型、清洗、統計、EDA/CDA、假設檢定	監督/非監督/強化、模型、Loss、評估、深度學習	DAI/GAI、GAN、VAE、Diffusion、整合應用

二、★★★★★ 必背 20 句

- 1. AI 是大範圍概念；機器學習與深度學習都是 AI 的重要技術。
- 2. 分析型 AI＝找洞察；預測型 AI＝推估未來；生成型 AI＝產生新內容。
- 3. 結構化＝固定行列；半結構化＝JSON/XML/CSV；非結構化＝圖片/音訊/文章。
- 4. 資料清洗常見問題：遺缺、重複、錯誤、不一致。
- 5. 離群值不一定是錯誤，是否刪除要看資料與業務情境。
- 6. Mean 易受極端值影響；Median 對極端值較不敏感；Mode＝出現最多。
- 7. SD＝分散程度；SD 越大，資料通常越分散。
- 8. IQR＝Q3－Q1＝中間 50% 資料範圍。
- 9. 正偏態：Mean>Median；負偏態：Mean<Median。
- 10. EDA＝探索；CDA＝驗證；Diagnostic＝為什麼；Predictive＝未來。
- 11. 監督式＝有 Label；非監督式＝無 Label；強化學習＝Reward。
- 12. Classification＝預測類別；Regression＝預測連續數值。
- 13. Linear Regression＝連續數值；Logistic Regression＝分類。
- 14. KNN＝看鄰居；Decision Tree＝樹狀規則；Random Forest＝多棵樹集成。
- 15. K-Means＝分群；PCA＝降維。
- 16. Loss＝衡量誤差；Gradient Descent＝調整參數以降低 Loss。
- 17. Overfitting＝訓練好、測試差；可用 Regularization、Early Stopping 等降低。
- 18. CNN＝圖像；RNN＝序列；LSTM＝長期依賴。
- 19. GAN＝Generator 生成＋Discriminator 判斷；常見問題 Mode Collapse。
- 20. 鑑別式 AI＝判斷/分類/預測；生成式 AI＝產生新資料。

三、模型一眼判斷表

技術	用途	例子	口訣
Linear Regression	連續數值	房價、銷售額	看到「多少錢／數值」先想它
Logistic Regression	分類	風險、是否購買、是否違約	看到「是/否、0/1」先想它
KNN	鄰近分類/迴歸	依相似樣本判斷	K=鄰居數
SVM	分類邊界	文本、影像分類	最大化類別間隔
Decision Tree	分類/預測	診斷、風險	樹狀規則、可解釋
Random Forest	集成分類/預測	分類、風險	很多棵 Decision Tree
K-Means	聚類	客戶分群	K=群數
PCA	降維	高維資料壓縮/視覺化	不是分類器
CNN	圖像	影像辨識、物件偵測	Convolution
RNN	序列	文字、語音、時間序列	前一時刻狀態
LSTM	長期序列依賴	文字、時間序列	RNN 改良、門機制
GAN	生成	影像生成、資料增強	Generator+Discriminator
VAE	生成/重建	合成資料、異常檢測	Encoder+Latent+Decoder
Diffusion	生成	高品質圖像、文字到圖像	加雜訊→反向去雜訊

四、資料分析：最容易出題的地方

① 資料型態

- 結構化：MySQL、PostgreSQL、規範化電子表格。
- 半結構化：JSON、XML、CSV。
- 非結構化：圖片、影像、音訊、電子郵件、文章。

② 資料清洗

- 遺缺值：平均數／中位數／眾數／插補／預測模型；也可能刪除，但須評估代表性。
- 重複值：依主鍵或唯一識別碼檢查。
- 錯誤值：例如年齡為 -5。
- 離群值：可能是真異常，也可能是錯誤或雜訊；不要看到離群值就直接刪。

③ 統計數字看到就要會

名稱	意思	記憶
Mean	平均數	容易受極端值影響
Median	中位數	排序後中間值
Mode	眾數	出現最多，可多個
SD	標準差	越大→越分散
IQR	$Q3 - Q1$	中間 50%

④ 分析方法

- EDA：先探索資料，不預設假設；找模式、關係、異常。
- CDA：驗證已提出的假設。
- 診斷性分析：回答「為什麼發生？」
- 預測性分析：回答「未來可能發生什麼？」

⑤ 假設檢定

- H_0 = 虛無假設； H_1/H_a = 對立假設。
- α 常見 0.1、0.05、0.01。
- $p < \alpha \rightarrow$ 拒絕 H_0 （依本講義的考試脈絡）。
- Type I Error： H_0 真卻拒絕 H_0 。
- Type II Error： H_a 真卻未拒絕 H_0 。

五、機器學習：考題看到關鍵字就選

題目關鍵字	答案方向	代表技術
有標記／有正確答案	監督式	Classification / Regression
沒有標記／找群組	非監督式	K-Means
降維／高維資料	非監督式／資料縮減	PCA
遊戲／獎勵／試錯	強化學習	Q-Learning
房價／銷售額／溫度	連續數值預測	Linear Regression
垃圾郵件／良惡性／是否違約	分類	Logistic / Tree / SVM
客戶分群	聚類	K-Means
影像辨識	深度學習／圖像	CNN
文字／時間序列	序列	RNN / LSTM
生成圖片	生成式	GAN / Diffusion

六、模型訓練與評估：一定要分清楚

- Loss Function：預測與實際的差異；Loss 越高代表誤差越大。
- MSE：本講義列為常用迴歸損失；Cross-Entropy：常用分類。
- Gradient Descent：調整模型參數，目標是降低 Loss。
- Overfitting：訓練表現佳、測試表現差。
- 降低 Overfitting：Regularization、Early Stopping、Data Augmentation。
- Accuracy：分類正確比例。
- F1：綜合 Precision 與 Recall；本講義指出適合不平衡資料。
- K-Fold Cross Validation：分 K 折輪流測試，最後平均結果。

七、深度學習與生成式 AI：最後一定要背

神經網路

- 基本結構：Input Layer → Hidden Layer(s) → Output Layer。
- Backpropagation：利用誤差回傳調整權重。

CNN / RNN / LSTM

- CNN：擅長圖像；卷積層擷取局部特徵；池化可降低維度。
- RNN：適合序列資料，利用前一時刻狀態。
- LSTM：RNN 改良，利用門機制改善長期依賴與梯度消失問題。

GAN：最常考的四件事

- Generator：生成資料。
- Discriminator：判斷真偽。
- 兩者透過對抗學習共同訓練。
- 常見問題：Mode Collapse。

VAE / Diffusion

- VAE：Encoder → Latent Space → Decoder。
- Diffusion：逐步加入雜訊，再透過反向過程逐步去除雜訊生成資料。

八、鑑別式 vs 生成式：10 秒判斷法

	鑑別式 AI	生成式 AI
一句話	判斷	創造
主要工作	分類、迴歸、辨識、預測	產生文字、圖片、語音等新內容
核心	學習判別關係	模擬資料分佈
典型	CNN、Logistic 等分類模型	GAN、VAE、Diffusion
例子	判斷腫瘤良惡性	生成病理影像

九、AI 治理：容易被忽略，但要背

- 六個核心字：透明、公平、隱私、安全、責任、倫理。
- 透明性：透過報告、技術文件或網站等方式揭露 AI 系統相關資訊。
- 公平性：避免因訓練資料偏差造成不同族群的不公平結果。
- 隱私：使用個人資料時要注意資料保護。
- 安全：降低惡意利用及系統錯誤造成的風險。
- 國際合作：共同標準、降低濫用風險、促進技術交流與技術轉移。
- Human-in-the-loop：重要決策／行動前由人類審核。
- Human-over-the-loop：系統自主運作，人類持續監督並必要時介入。

十、最容易混淆的 15 組

易混淆

一句話拆開

AI vs Deep Learning

AI 是大領域；深度學習是機器學習的子領域。

分析型 vs 預測型

分析現在/過去資料找洞察；預測未來。

預測型 vs 生成型

預測既有目標；生成新內容。

結構化 vs 半結構化

結構化固定行列；半結構化有標籤但彈性較高。

EDA vs CDA

EDA 探索；CDA 驗證假設。

診斷 vs 預測

診斷問為什麼；預測問未來。

監督 vs 非監督

有 Label；無 Label。

分類 vs 迴歸

類別；連續數值。

K-Means vs KNN

K-Means 分群；KNN 看鄰居分類/迴歸。

Tree vs Forest

一棵樹；多棵樹集成。

CNN vs RNN

圖像；序列。

RNN vs LSTM

一般序列；更強調長期依賴。

GAN Generator vs Discriminator

生成；判真假。

鑑別式 vs 生成式

判斷；創造。

Type I vs Type II

H0 真卻拒絕；Ha 真卻未拒絕。

十一、考前 30 分鐘衝刺法

- 第 1~5 分鐘：背「必背 20 句」。
- 第 6~10 分鐘：背模型表：Linear / Logistic / KNN / Tree / Forest / K-Means / PCA。
- 第 11~15 分鐘：背 CNN / RNN / LSTM / GAN / VAE / Diffusion。
- 第 16~20 分鐘：重看資料清洗、Mean/Median/Mode/SD/IQR、EDA/CDA。
- 第 21~25 分鐘：重看監督／非監督／強化與 Classification／Regression。
- 第 26~30 分鐘：只看「最容易混淆的 15 組」。

十二、最後 1 頁速查卡

- AI=大範圍；ML、DL=AI 技術。
- 分析=洞察；預測=未來；生成=創造。
- Structured=表格；Semi=JSON/XML/CSV；Unstructured=圖片/音訊/文章。
- Clean：Missing / Duplicate / Error / Inconsistency。
- Mean 易受極端值；Median 不敏感；Mode 最多；SD 分散；IQR=Q3-Q1。
- 正偏：Mean>Median；負偏：Mean<Median。
- EDA 探索；CDA 驗證；Diagnostic 為什麼；Predictive 未來。
- Supervised=Label；Unsupervised=無 Label；RL=Reward。
- Classification=類別；Regression=連續數值。
- Linear=數值；Logistic=分類；KNN=鄰居；Tree=樹；Forest=多樹。
- K-Means=分群；PCA=降維。
- Loss=誤差；Gradient Descent=降 Loss；CV=穩健評估。
- Overfit=Train 好 Test 差。
- CNN=圖像；RNN=序列；LSTM=長期依賴。
- GAN=G 生成+D 判斷；VAE=Encoder+Decoder；Diffusion=加噪→去噪。
- 鑑別式=判斷；生成式=創造。
- 治理六字：透明、公平、隱私、安全、責任、倫理。

考試策略：先找關鍵字，再判斷類型，最後才看選項細節。