# 📘 第九章：人工神經網路與深度學習（Artificial Neural Networks and Deep Learning）

## 📖 章節簡介與補充重點

本章介紹人工神經網路（ANN）的基本架構與運作原理，包括感知機、多層感知機（MLP）、啟動函數、損失函數與反向傳播法（Backpropagation）。此外也探討深度學習（Deep Learning）的優勢、挑戰與常見問題（如梯度消失）。  
  
重點主題包括：  
1. 感知機與 MLP 的架構與數學模型。  
2. 常見啟動函數（Sigmoid、Tanh、ReLU）與特性比較。  
3. 損失函數與反向傳播（含梯度計算與參數更新）。  
4. 深度學習中的困境與改善方法（如 Batch Normalization）。

## 🧠 名詞解釋（Term Explanation）

感知機（Perceptron）  
中文：最早期的神經元模型，透過加權輸入與閾值判斷輸出。  
英文：The earliest neural model that uses weighted inputs and a threshold to decide output.

多層感知機（MLP）  
中文：具有一個或多個隱藏層的神經網路，可處理非線性問題。  
英文：Neural networks with one or more hidden layers capable of modeling nonlinear patterns.

啟動函數（Activation Function）  
中文：決定神經元輸出的非線性函數，如 Sigmoid、ReLU。  
英文：Nonlinear functions like Sigmoid or ReLU that determine neuron output.

ReLU 函數  
中文：常用啟動函數，輸入大於 0 則輸出本身，否則為 0。  
英文：A popular activation function: output equals input if positive, else 0.

梯度消失問題  
中文：在深層網路中，反向傳播時梯度逐漸趨近於 0，導致前層無法有效學習。  
英文：In deep networks, gradients shrink during backpropagation, hindering learning in early layers.

反向傳播法（Backpropagation）  
中文：透過鏈式法則計算損失函數對每層參數的梯度，用以更新權重。  
英文：Computes gradients layer-by-layer via chain rule to update weights.

損失函數（Loss Function）  
中文：衡量模型預測與實際目標之間的誤差，如 MSE、Cross-Entropy。  
英文：Measures prediction error, e.g., Mean Squared Error or Cross-Entropy.

深度學習（Deep Learning）  
中文：利用多層神經網路學習高層特徵表示的技術。  
英文：Uses multi-layer neural networks to learn hierarchical feature representations.

Batch Normalization  
中文：在每層輸出上進行正規化，加速訓練與穩定梯度。  
英文：Normalizes intermediate layer outputs to accelerate training and stabilize gradients.

Dropout  
中文：隨機忽略部分神經元，以減少過擬合風險。  
英文：Randomly drops neurons during training to reduce overfitting.

## 📚 名詞比較（Term Comparison）

📌 Sigmoid vs. ReLU vs. Tanh  
中文：Sigmoid 與 Tanh 在輸入極端時容易梯度消失；ReLU 可避免此問題但可能造成 dead neurons。  
英文：Sigmoid/Tanh may cause vanishing gradients; ReLU avoids that but risks dead neurons.  
  
📌 MLP vs. Logistic Regression  
中文：Logistic Regression 是單層線性分類器；MLP 含隱藏層可處理非線性問題。  
英文：Logistic Regression is a single-layer linear classifier; MLP handles nonlinear patterns with hidden layers.  
  
📌 Backpropagation vs. Gradient Descent  
中文：Backpropagation 計算梯度；Gradient Descent 使用梯度來更新參數。  
英文：Backpropagation computes gradients; Gradient Descent updates parameters using those gradients.

## 📝 模擬出題與中英文詳解（包含解釋性問題）

Q1: Q1. 為什麼 MLP 能處理非線性分類，而感知機無法？請解釋其差異。  
答：感知機僅能處理線性可分問題；MLP 透過隱藏層與非線性啟動函數能擬合複雜邊界。  
Perceptron handles only linearly separable data; MLP can model complex boundaries using hidden layers and nonlinear activations.

Q2: Q2. 為何深度神經網路會出現梯度消失問題？有何解法？  
答：因為 Sigmoid/Tanh 的梯度會隨層數遞減；可用 ReLU 或 Batch Normalization 改善。  
Due to shrinking gradients in Sigmoid/Tanh; solved by ReLU or Batch Normalization.

Q3: Q3. 請說明反向傳播法如何計算梯度？為何使用鏈式法則？  
答：反向傳播利用鏈式法則，逐層從輸出往前計算誤差對權重的偏導數。  
Backpropagation uses chain rule to compute weight gradients layer by layer from output backward.

Q4: Q4. 在訓練深度學習模型時，Dropout 的目的為何？其如何改善模型泛化能力？  
答：Dropout 隨機忽略神經元，減少過擬合，讓模型不依賴特定神經元。  
Dropout randomly disables neurons, reducing overfitting and encouraging robustness.