雲端計算實驗課程　助教 楊哲旻

1. 為何資料要分為訓練集與測試(驗證)集？

因為大多複雜的模型容易擬合訓練集的資料，導致實際上應用預測能力沒那麼高，為了避免這情況發生，因此會將資料分為兩部分，其中訓練集當作模型訓練用，而測試(驗證)集作為驗證預測結果。

比如：平常練習課本的練習題 與 期中期末考

2. 如何「公平地」分為訓練集與測試集？

資料集在訓練集與測試集的標籤**類別比例要均衡**。原因是不均衡時，模型容易傾向於學習標籤類別數量多的，也原在於損失函數的設計。

回歸 $Loss\ function_{\ MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y^{(i)} - y^{(i)})^2$

分類 $Loss\ function_{\ Cross-Entropy} = -\sum_{i=1}^{n} Y^{(i)} log y^{(i)} + (1 - Y^{(i)})\log(1 - y^{(i)})$

比如：平常學生九成讀數學都花在作三角函數題目，因此考試時只會三角函數，其餘像是排列組合都不太會算。

3. 為何模型訓練時會需要定義損失函數？它的功能為？

目的是為了想讓模型知道，訓練集中這某個(些)資料預測的好壞，從預測值與實際值(標籤)去比對，讓模型能從誤差中再一次更新權重，使決策函數(模型)朝向誤差越小的狀態。

回歸 $Loss\ function_{\ MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y^{(i)} - y^{(i)})^2$

分類 $Loss\ function_{\ Cross-Entropy} = -\sum_{i=1}^{n} Y^{(i)} log y^{(i)} + (1 - Y^{(i)})\log(1 - y^{(i)})$

4. 如果資料真的不足或是標籤類別數量不均勻，該如何解決？

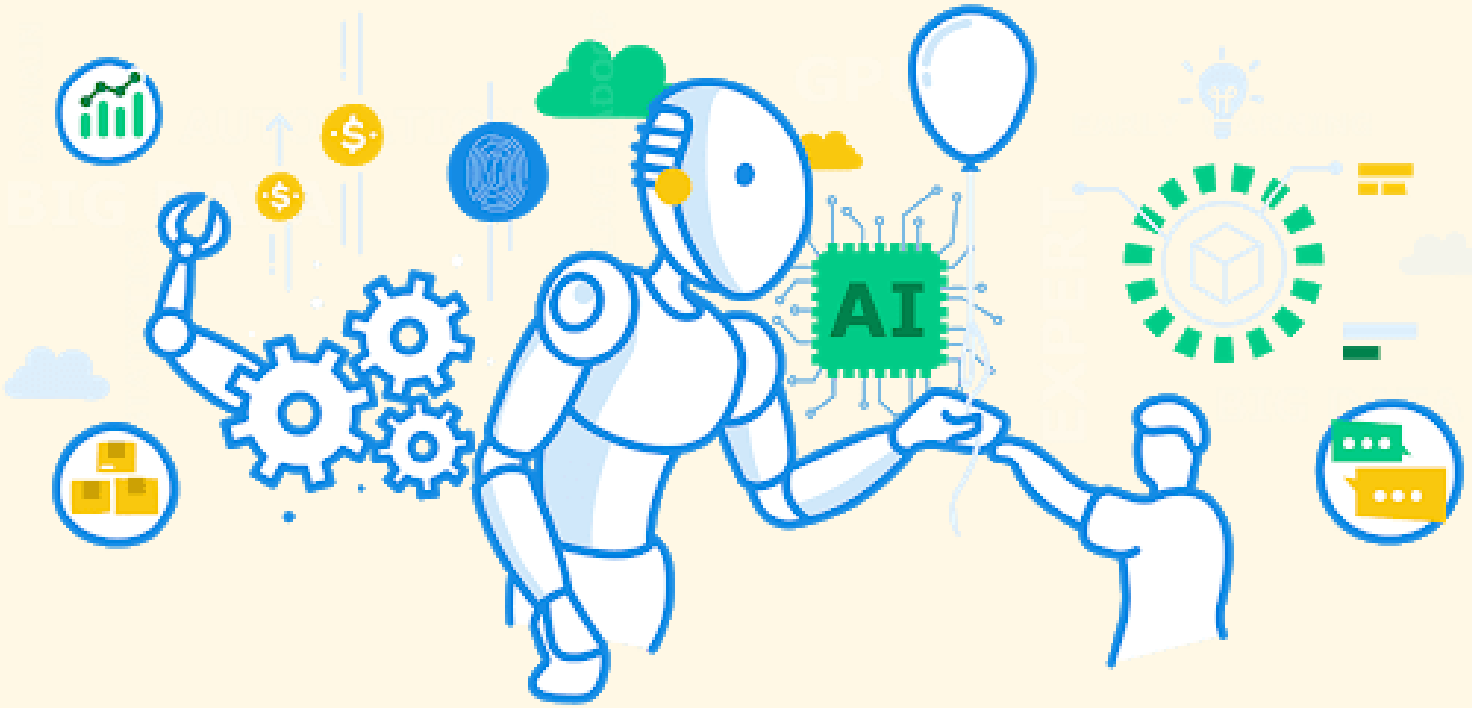
解決方式：

1. 收集更多資料
2. 資料前處理(影像增量：影像水平、翻轉、裁減等)

比如：

1. 多買點教科書與考卷
2. 題目不多，那就改數字，變成類似題目但還是稍微不同

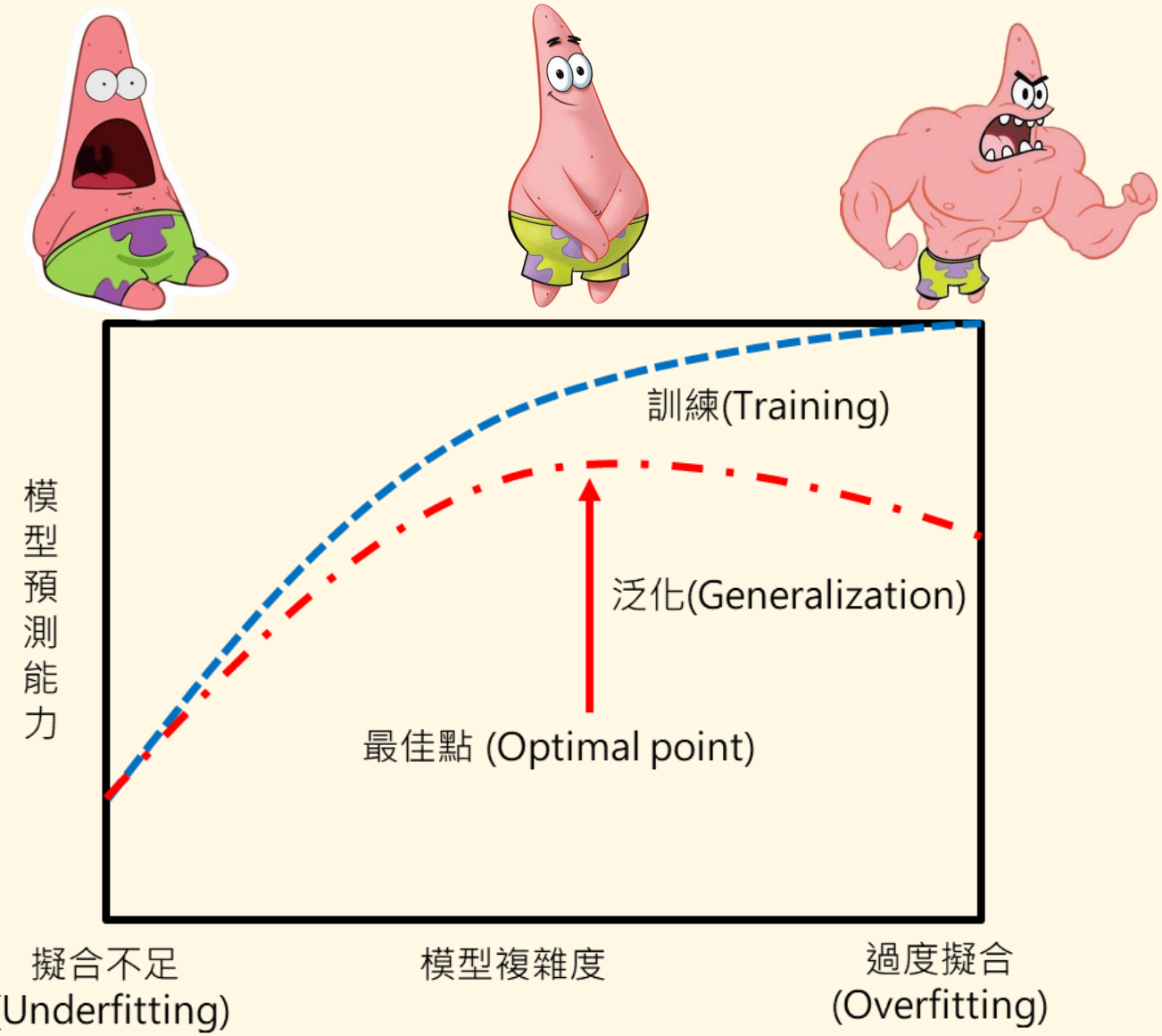
AI 複習



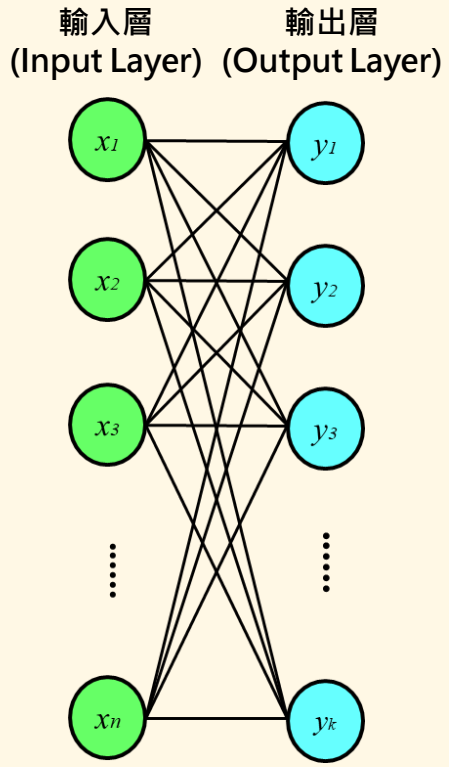
雲端計算實驗課程　助教 楊哲旻

1. 泛化能力(Generalization Ability)

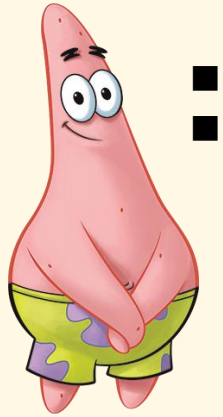
1. 資料樣本數量增大
2. 資料前處理
3. 調降低整模型**超參數**來模型複雜度



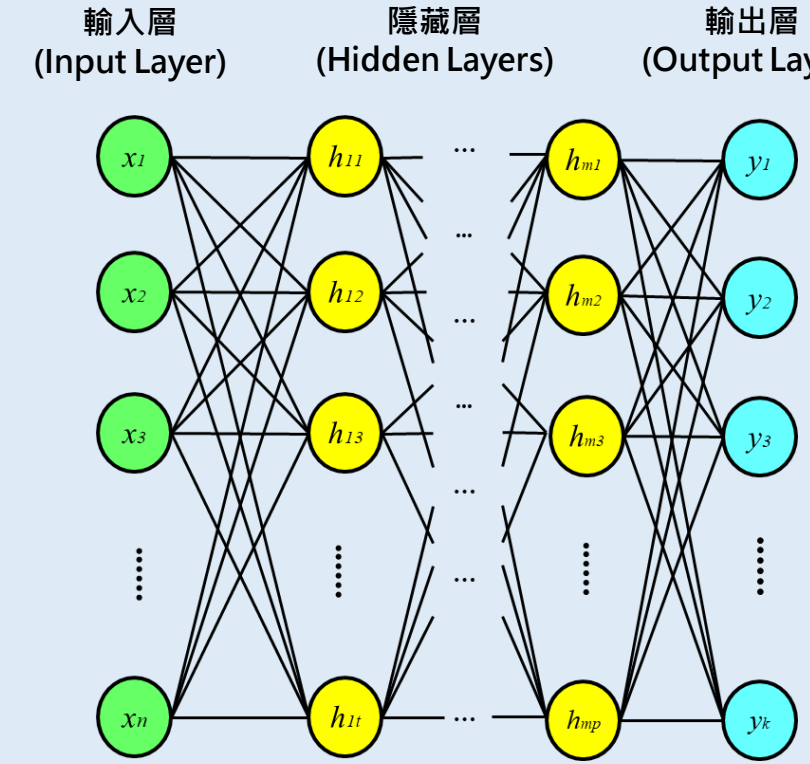
2. 分類模型：邏輯回歸、多層感知器、卷積神經網路



邏輯回歸

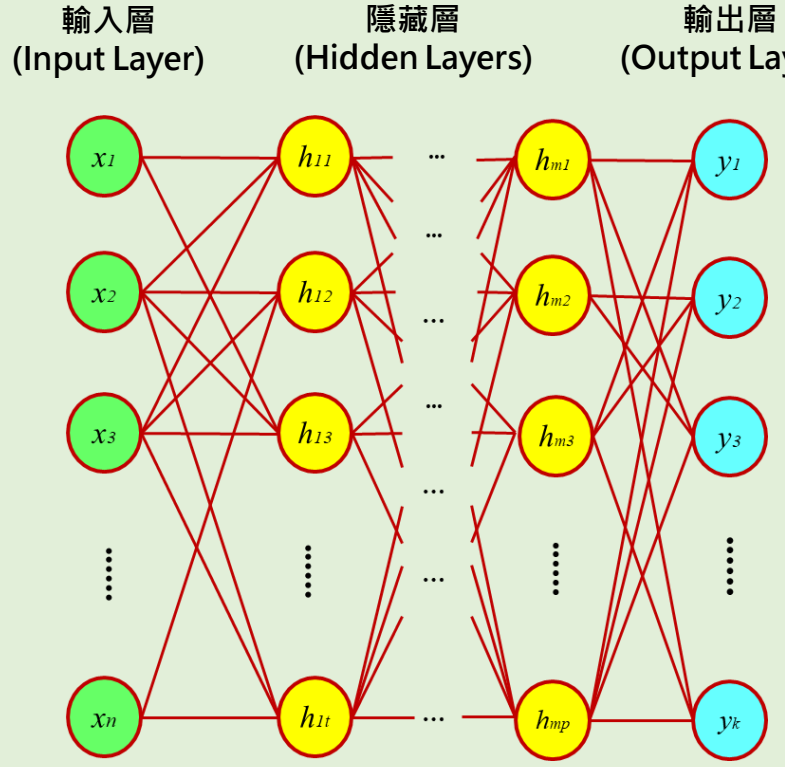


- 屬機器學習
- 單層



多層感知器

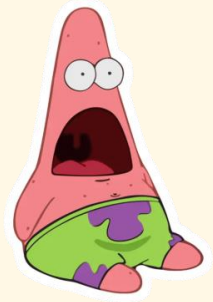
- 屬深度學習
- 多層



卷積神經網路

- 屬深度學習
- 多層
- 用於影像分類，神經元非完全連接，考慮影像局部特徵)

擬合不足　　最佳　　過度擬合



調整超參數

調整超參數

邏輯回歸

1. 批量
2. 學習速率
3. 迭代次數
4. 正則化係數

調整超參數

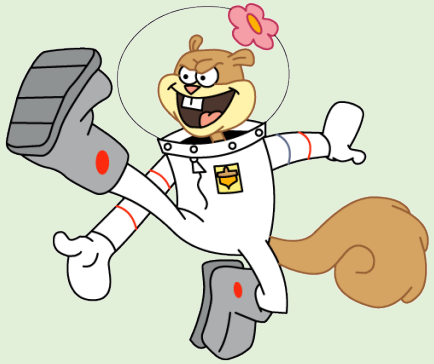
調整超參數

多層感知器

1. 批量
2. 學習速率
3. 迭代次數
4. 正則化係數
5. 隱藏層層數與神經元數量
6. 丟棄法

調整超參數

調整超參數



卷積神經網路

1. 批量
2. 學習速率
3. 迭代次數
4. 正則化係數
5. 隱藏層層數與神經元數量
6. 丟棄法
7. 卷積與池化層的濾波器大小與數量

3. 一維、二維資料分類

一維 K個特徵 資料

身高	體重	…	血壓	標籤
180	80	…	90	0
175	65	…	70	0
…	…	…	…	…
168	59	…	60	1

第一筆資料的第一個特徵：[180]	形狀：(1)
第一筆資料：[180 80 … 90]	形狀：(K)
N筆資料：[[180 80 … 90] [175 65 … 70] ⋮ ⋮ ⋮ [168 59 … 60]]	形狀：(N,K)

二維資料

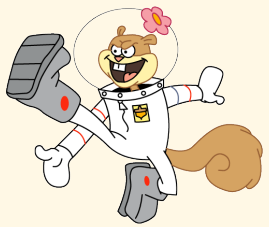
影像	標籤
	0
	0
	1
	1
…	…

4. 灰階的影像

Gray

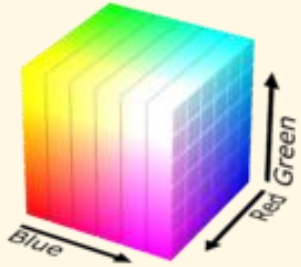
255	250	...	255
254	255	...	255
145	100	...	180
...	...	...	...
255	255	...	255

第一個像素：[255]	形狀：(1)
第一排的像素：[[255] [250] ... [255]]	形狀：(150,1)
一張圖的像素：[[[255] [250] ... [255]] [[254] [255] ... [255]] ⋮ ⋮ ⋮ [[255] [255] ... [255]]]	形狀：(120,150,1)
N張圖的像素	形狀：(N,120,150,1)



模型訓練

彩色的影像



哥哥! 彩色的我
(寬高150×120)
被拆成RGB三種
色塊的像素了

R

255	250	...	255
254	255	...	255
145	100	...	180
...	...	...	...
255	255	...	255

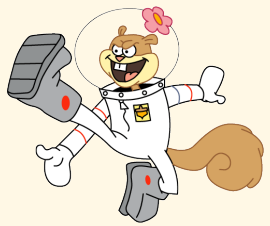
G

255	254	...	255
254	245	...	245
155	120	...	120
...	...	...	...
255	255	...	255

B

255	255	...	253
254	245	...	245
155	120	...	120
...	...	...	...
255	255	...	255

第一個像素：[255 255 255]	形狀：(3)
第一排的像素：[[255 255 255] [250 254 255] ... [255 255 253]]	形狀：(150,3)
一張圖的像素：[[[255 255 255] [250 254 255] ... [255 255 253]] [[254 254 254] [255 245 245] ... [255 245 245]] ⋮ ⋮ ⋮ [[255 255 255] [255 255 255] ... [255 255 255]]]	形狀：(120,150,3)
N張圖的像素	形狀：(N,120,150,3)



模型訓練

訓練前的四樣必備物品

1. N筆資料的特徵值
2. N筆資料的標籤值
3. 標籤的原始名稱
4. 特徵的原始名稱(非必要)