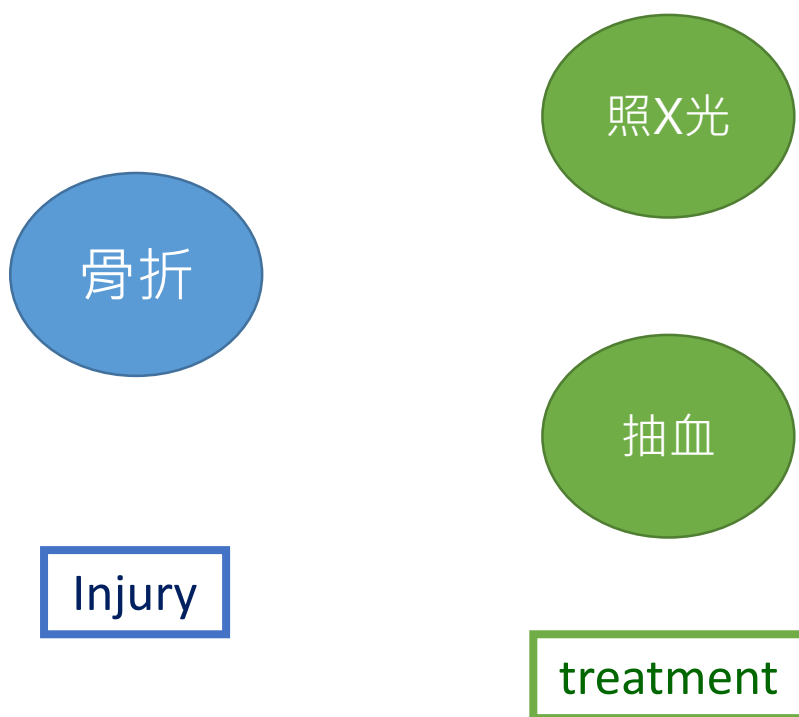


Deep Learning to Detect Medical Treatment Fraud

RBM如何有效地檢查醫療資源浪費

醫療浪費的問題

- 醫療資源浪費和濫用：最常見的問題是對患者進行過度治療或檢測
- 佔醫療支出的四分之一以上
- 佔美國衛生總支出的8%

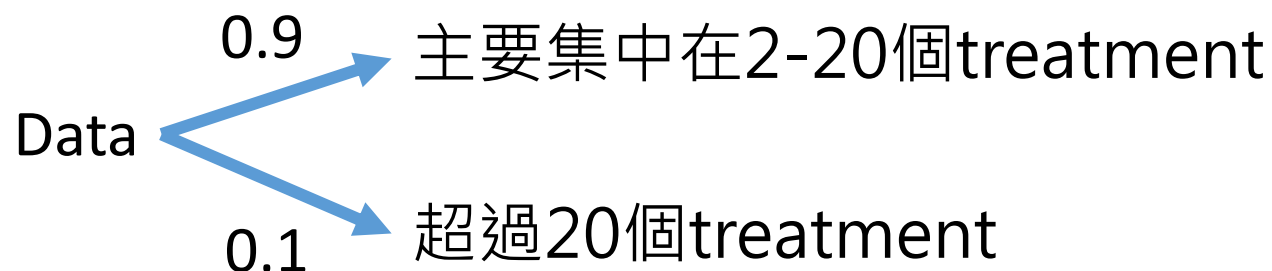


解決方法：

模擬Injury和treatment可能的組合，
利用RBM尋找其中異常(不必要)的治療。

數據模擬

- 建立一個矩陣 (y_t)
- 列：800種治療方式、行：200,000種受傷狀態
- 有10%異常的資料 (隨機選取)
- 把資料拆成90/10的training/testing data



要找出異常值

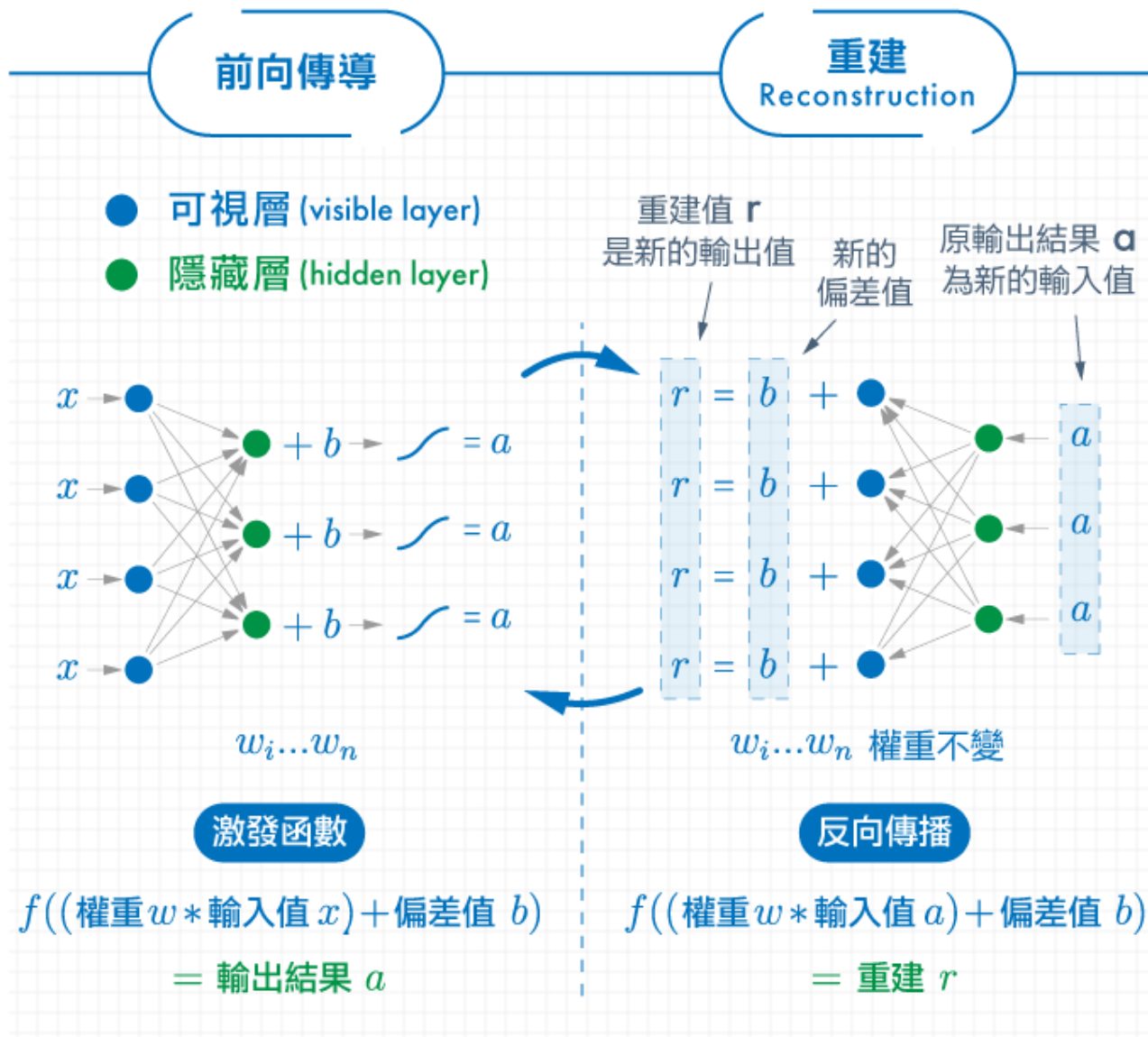
	x-ray	dilation	CAT-scan	Antibiotics
Injury 1	0	0	1	0
Injury 2	1	0	0	0
Injury 3	0	1	0	1
Injury 4	0	0	1	0

Table 1: Treatment occurrence matrix

RBM

- 兩層結構(可視層、隱藏層)的淺層神經網路
- 無監督學習-沒有輸出
- 目的：更新權重，讓預測出來的結果(\hat{y}_t)接近原始的資料(y_t)。
- 在過程中，尋找各類別之間所有潛在的關係。

RBM 結構



Injury Risk (MSE)

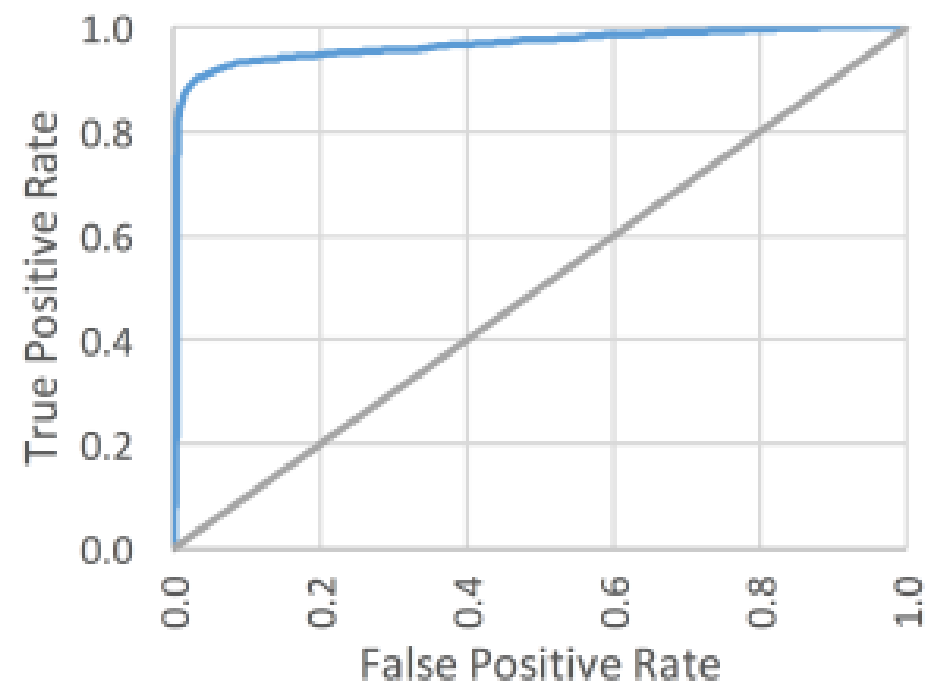
$$injury\ risk = \frac{\sum_t^T (y_t - \hat{y}_t)^2}{T}$$

- T : 治療類別數
- \hat{y}_t : 預測的組合
- y_t : 實際的組合

- Injury risk小 → 原始資料與預測的匹配(差異很小) → 正常治療檢測
- Injury risk大 → 原始資料與預測差異很大 → 異常治療檢測

挑最佳超參數

- Deepnet(R的套件)
- 固定→ sigmoid function、batch size=100、learning rate=0.8
- 調整→ hidden layer神經元個數、迭代次數
- 兩者的組合代入模型，用ROC/AUC來判斷模型的預測能力
- 最佳超參數→
 - 迭代次數=8
 - hidden layer神經元個數=1000
 - (AUC=0.96)



結論

- 矩陣中大多都是0，讓risk score介於0~0.003之間
- 調整→

$$rebalanced_score = \left(\frac{mse - \min(mse)}{\max(mse) - \min(mse)} \right)^{\frac{1}{9.85}}$$

- 依照調整過後的risk score畫直方圖，可明顯分辨出有無異常治療的差別。

