

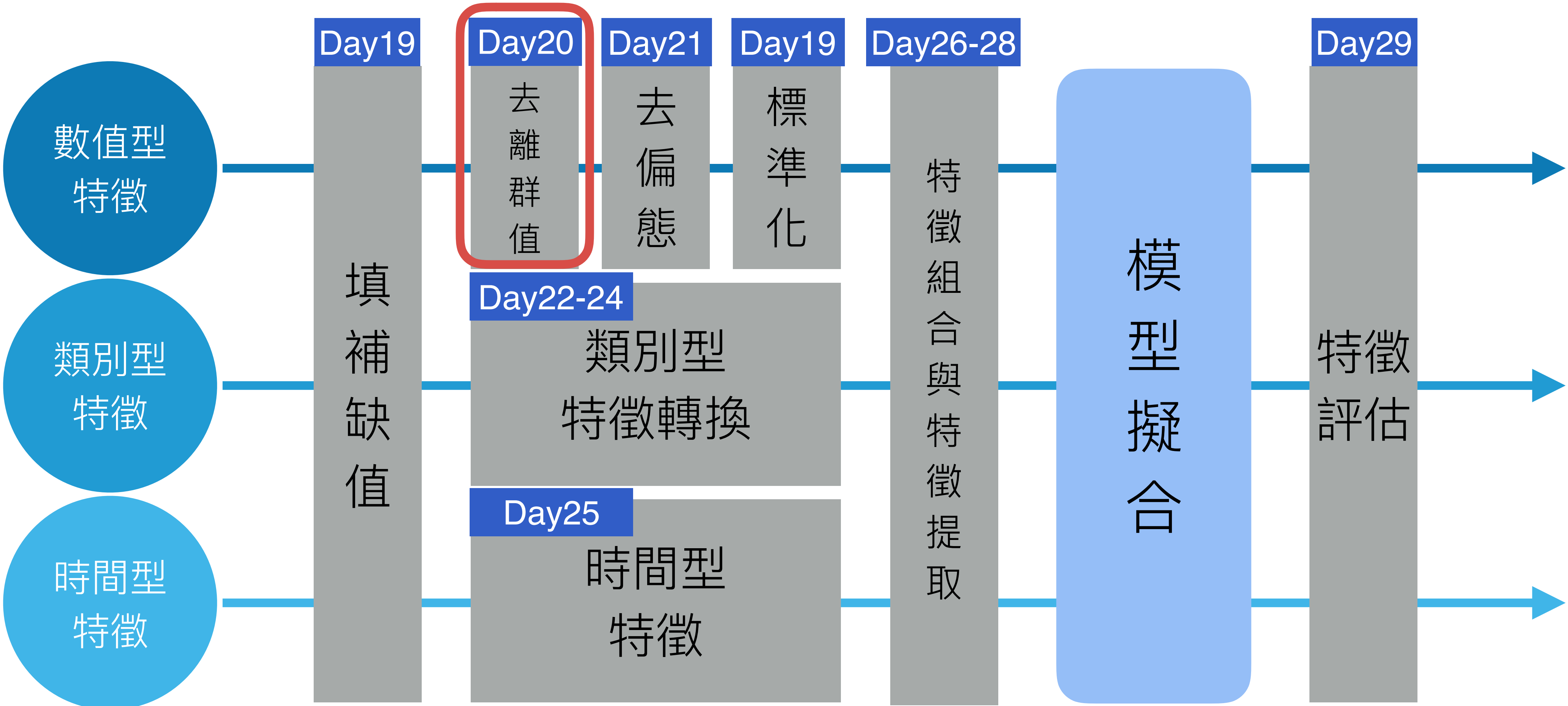
Day 20

特徵工程

數值型特徵-去除離群值

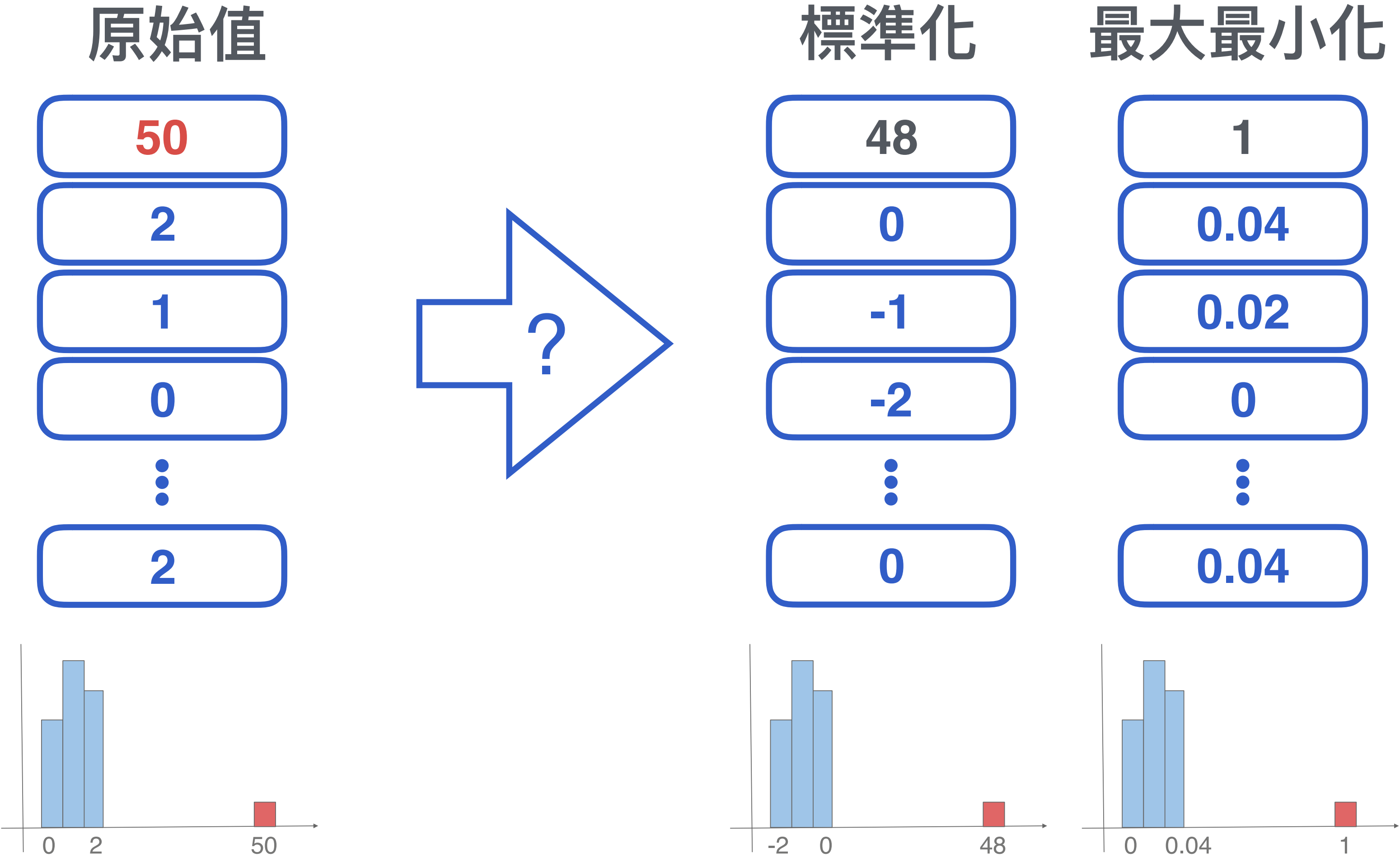


特徵工程 - 學習地圖



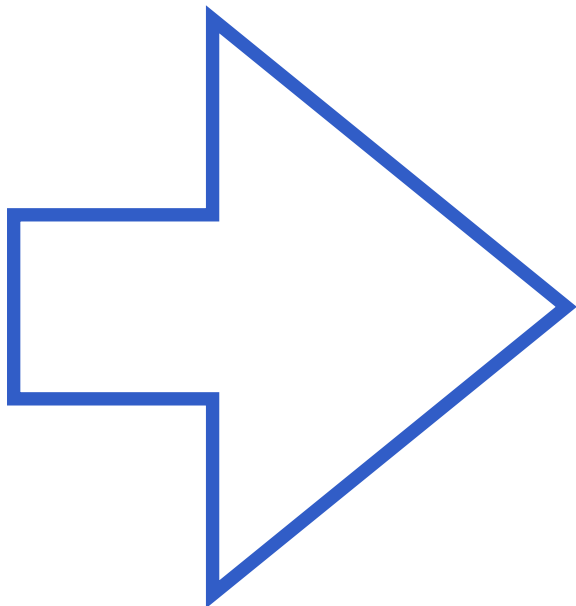
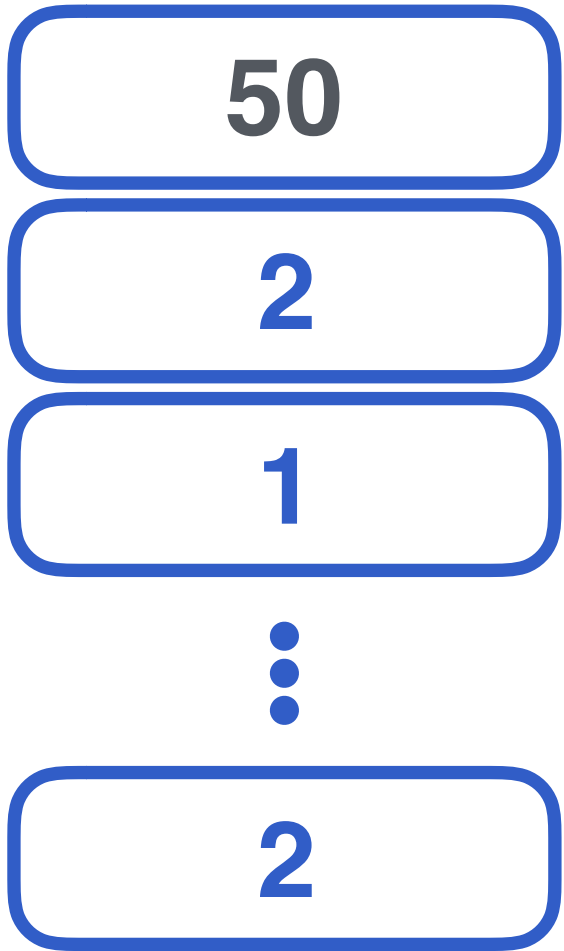
去除離群值 (1 / 2)

如果只有少數幾筆資料跟其他數值差異很大，標準化無法處理

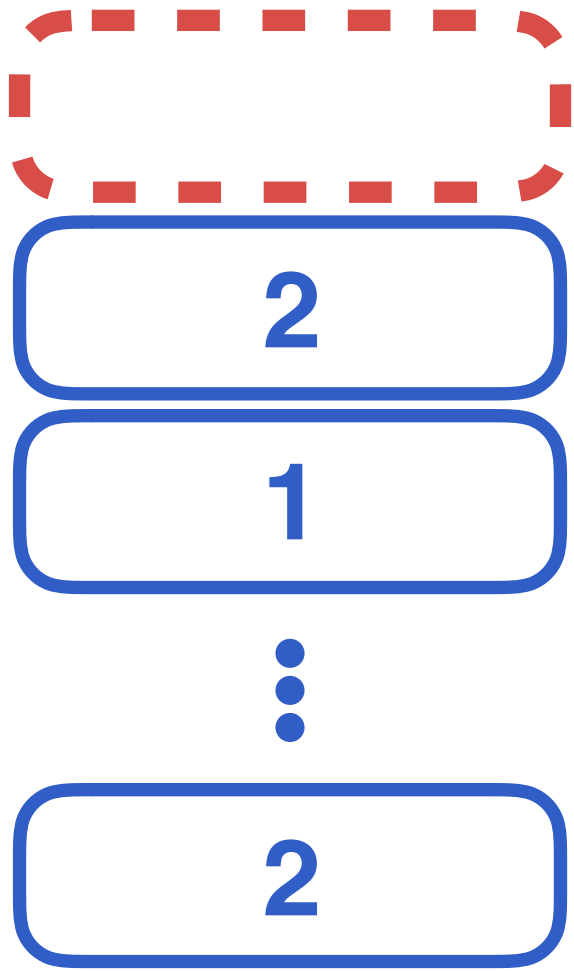


去除離群值 (2 / 2)

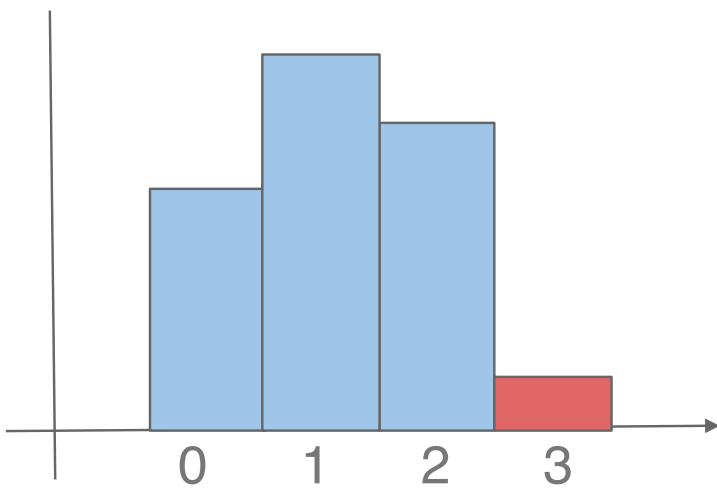
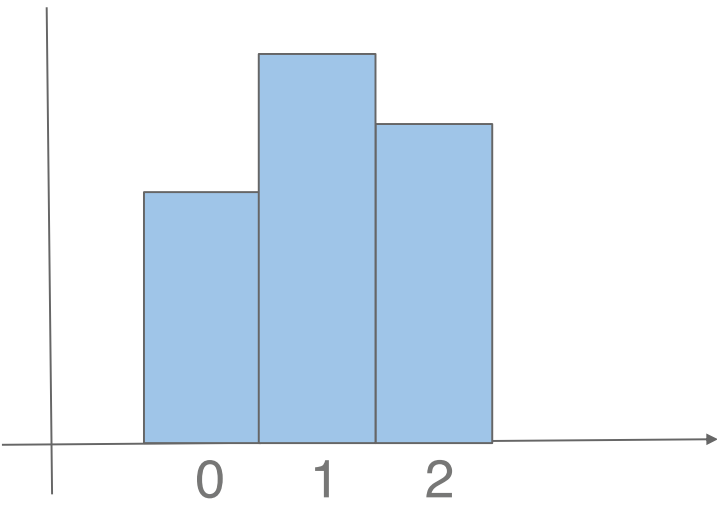
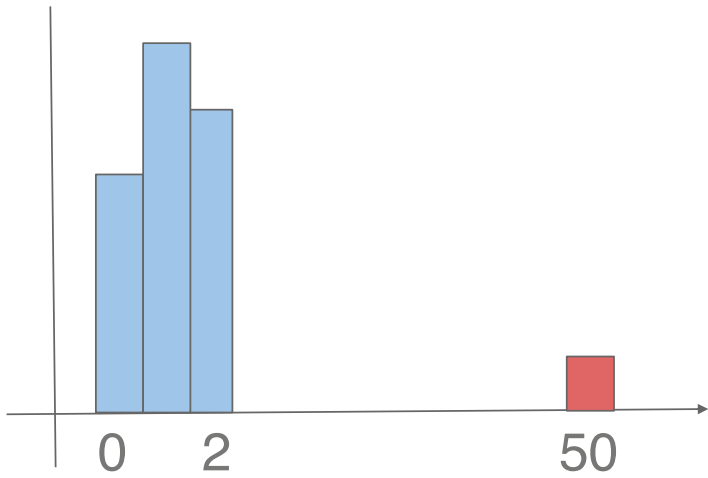
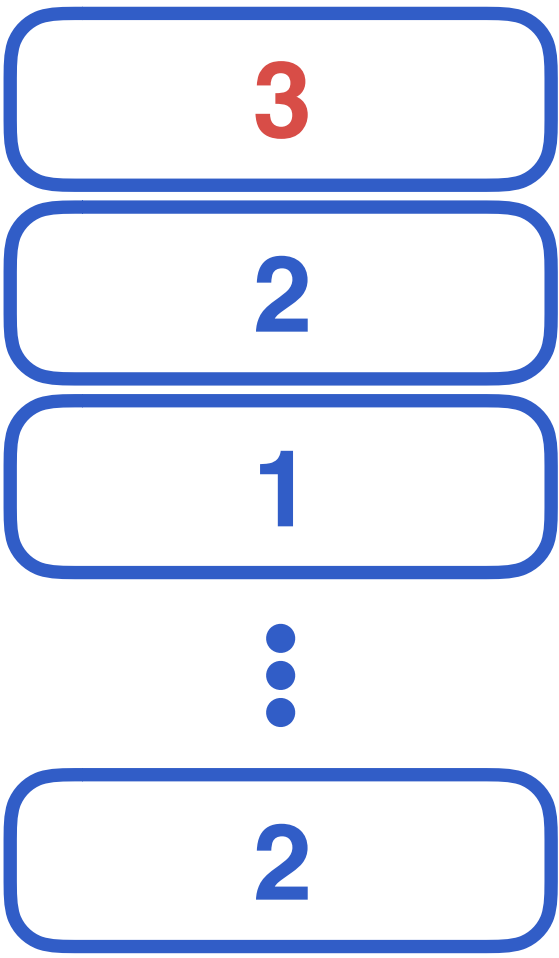
原始值



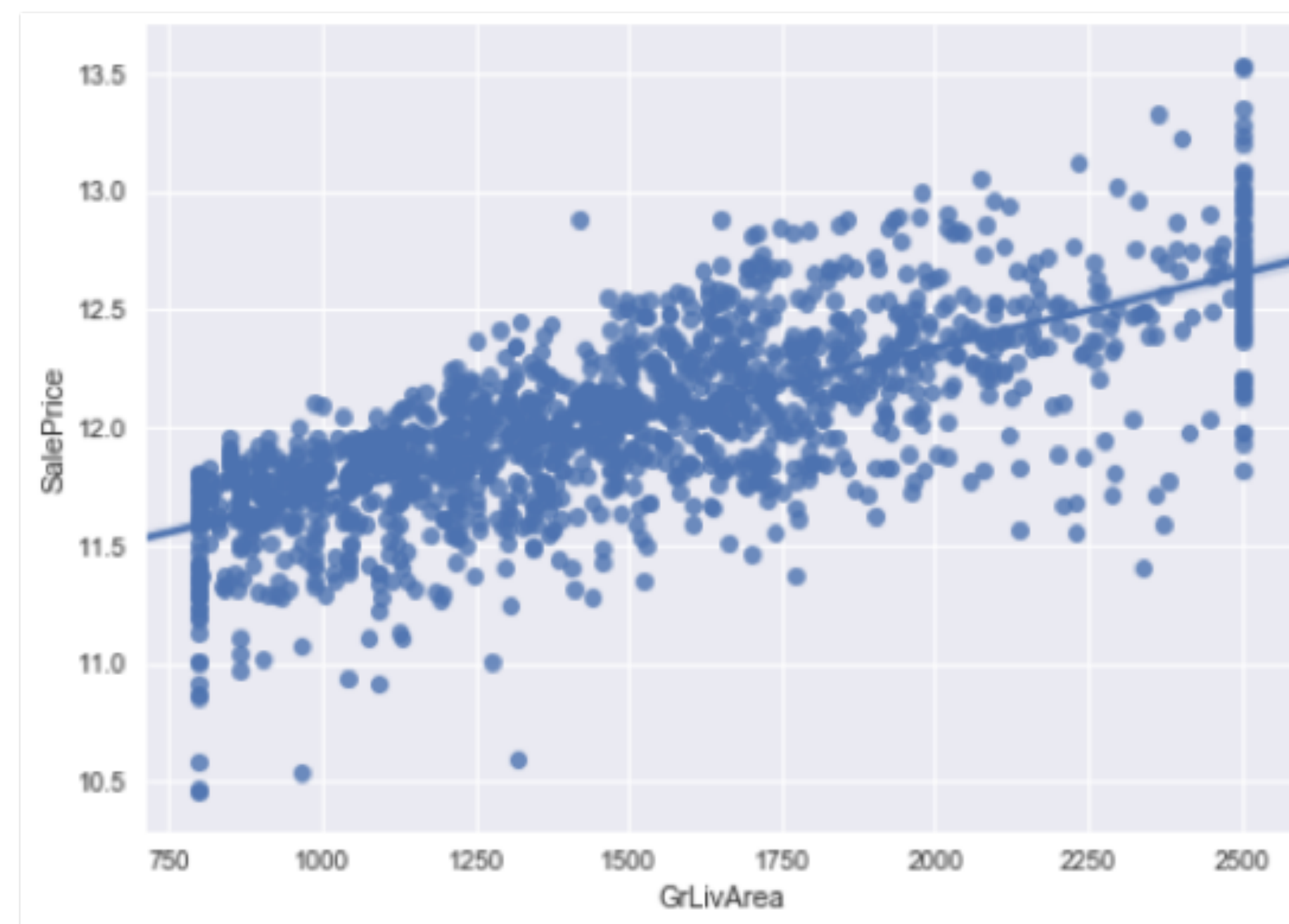
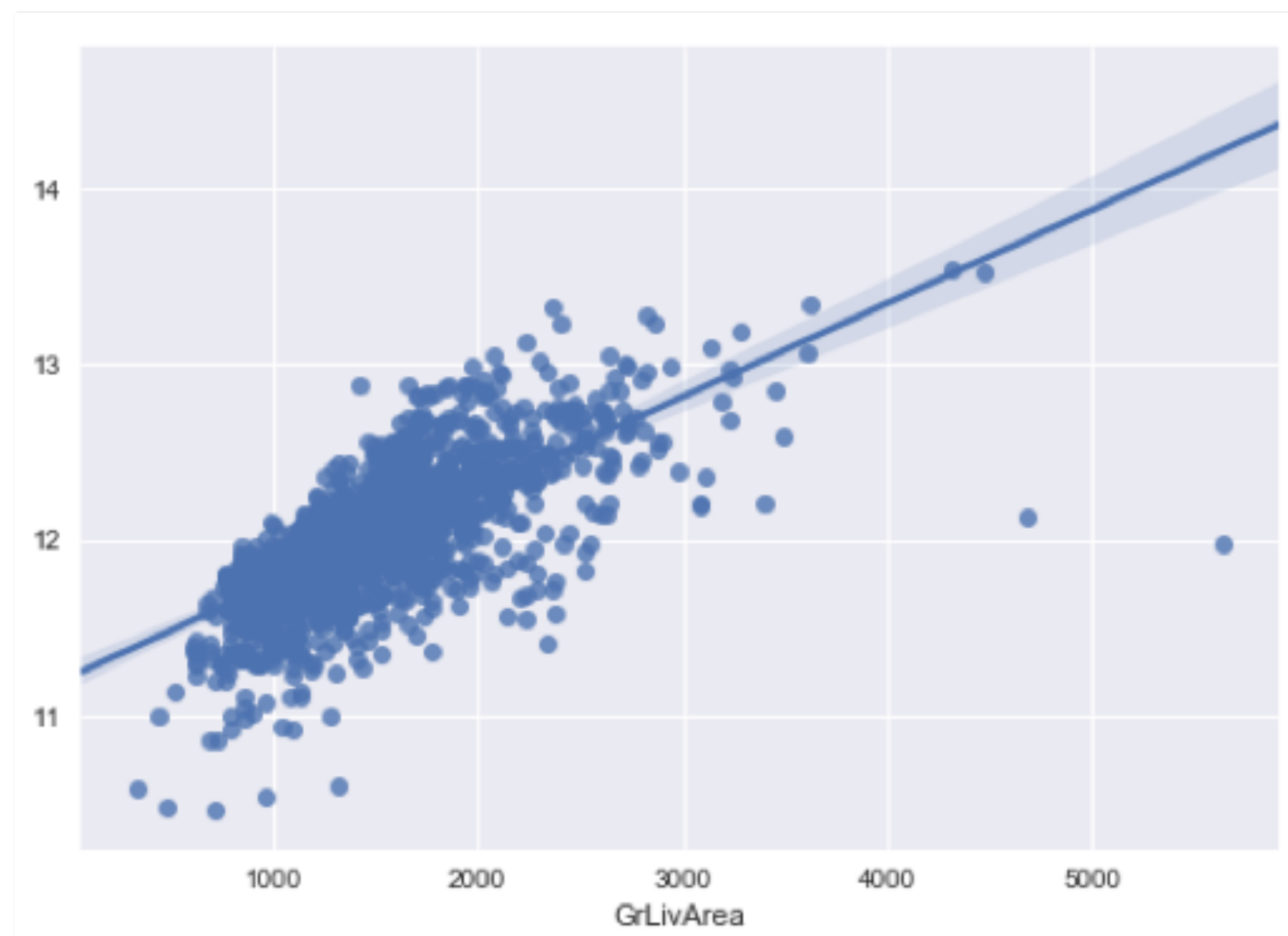
方法一
捨棄離群值



方法二
調整離群值



複習：去除離群值



數值型欄位，有時會有與其他數值差距很大的離群值存在
只要離群值數量夠少，除去離群值，將可使得模型預測較為準確
(參考圖中的迴歸直線以及今日範例)

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

