人工智慧導論

Homework 4

系所 : 資訊所

姓名 : 張少鈞

學號 : P76114545

# k-nearest-neighbors linear regression

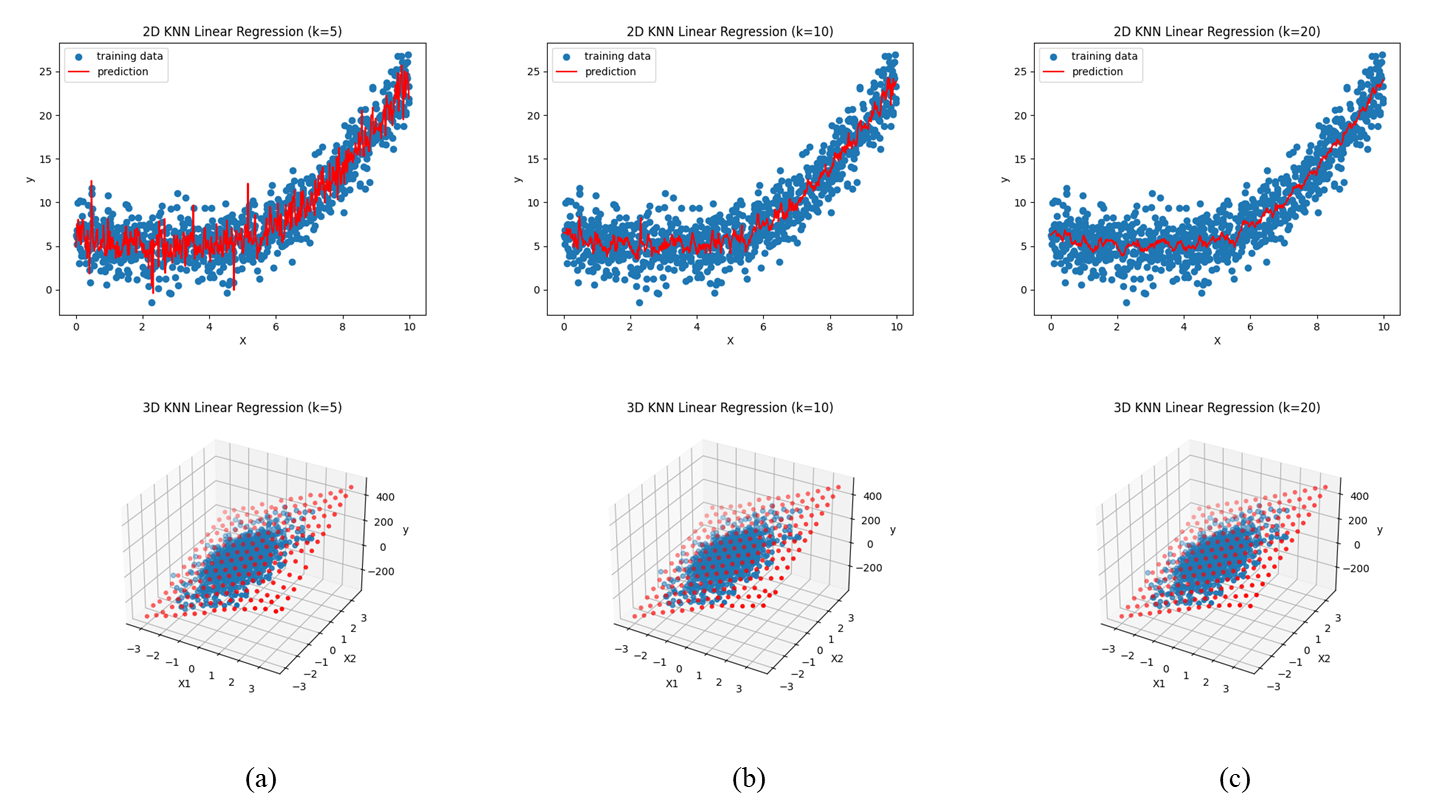
## Method

在進行n維的線性回歸問題，k-nearest-neighbors linear regression利用KNN的概念從訓練資料()中利用歐式距離(式 1)計算離查詢點()最接近的k個點。取出k個點()後利用最小平方法進行線性迴歸分析，最後回推查詢點所對應的輸出數值。

|  |  |
| --- | --- |
|  | 式 1 |

## Result

k-nearest-neighbors linear regression的結果如圖表 1。觀察結果可以發現，隨著k值的增加，KNN linear regression預測出來的數值越穩定，較不容易產生震盪，主要原因是k值的增加代表參考點增加，預測比較不會受到雜訊的干擾。



圖表 1 KNN linear regression result。(a) k=5 (b) k=10 (c) k=20

# Locally weighted regression

## Method

Locally weighted regression(LWR)透過給予不同的權重來改進傳統線性回歸模型欠擬合與過度擬合的問題，甚至能使用LWR來擬合非線性問題。

LWR的權重會依據查詢點而不同，越靠近查詢點的訓練資料權重會越高，反之則越低。第筆訓練資料的權重計算方式如(式 2)。式中的代表查詢點與各點的距離，而則代表權重隨距離下降的速度，最後透過的計算將限定在0~1之間。

|  |  |
| --- | --- |
|  | 式 2 |

在LWR的假設中其Loss函式定義如式 3，透過計算出各筆訓練資料的權重後，解出能使Loss降到最低的值，最後用此值來計算出查詢點的對應數值(式 4)。

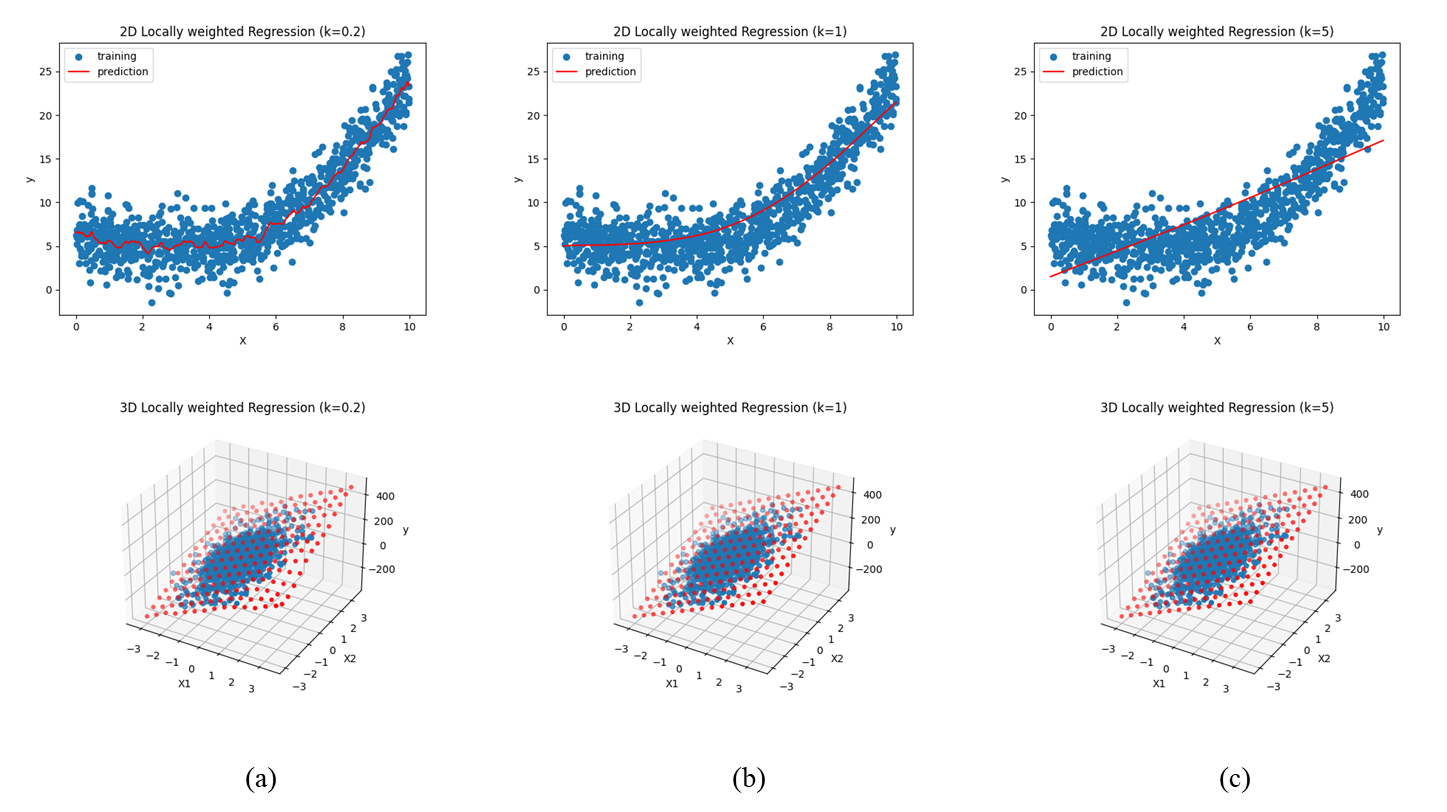
|  |  |
| --- | --- |
|  | 式 3 |
|  | 式 4 |

在求解的問題上直接使用矩陣計算來進行，將假設為0進行計算，最後將方程式改寫為式 5，即可計算出相對應的值。

|  |  |
| --- | --- |
|  | 式 5 |

## Result

透過觀察LWR的結果(圖表 2)發現，在k = 5時LWR的預測結果與線性回歸的結果相近，從LWR的權重公式來看，當k足夠大時每筆訓練資料的權重都會趨近於1，也就會變成傳統的線性迴歸分析問題。因此隨著k值的縮小，距離查詢點越遠的訓練點的權重會隨之變小，使預測更接近非線性的迴歸分析，因此震盪也隨之增加。



圖表 2 Locally weighted regression result。(a) k=0.2 (b) k=1 (c) k=5

# k-nearest-neighbors regression

## Method

k-nearest-neighbors regression是k-nearest-neighbors linear regression的前身，兩者之間的差別在於，從訓練資料取出k個點後，k-nearest-neighbors linear regression是使用這k筆資料進行線性迴歸分析來求查詢點的對應數值；反之，k-nearest-neighbors regression是將這k個點的對應值()取平均後當成查詢點的對應數值(如式 6)。

|  |  |
| --- | --- |
|  | 式 6 |

## Result

k-nearest-neighbors regression 為k-nearest-neighbors linear regression的前身，因此將兩者結果進行比較。透過觀察k-nearest-neighbors regression的結果可以發現，在相同的k值下k-nearest-neighbors regression的預測震盪結果會較小，因為其採用平均的方法來進行預測，而k-nearest-neighbors linear regression因為採用最小平方法計算線性回歸，容易造成預測之回歸直線斜率變化大，因此其預測結果也較容易產生震盪。

