**Introduction to Artificial Intelligence - Homework 4**

NE6114011 人工智慧所碩一 楊雲翔

**1 K-nearest-neighbors linear regression**

**實現方法**

存在訓練資料(X, Y)，其中X=()，Y)，，n表資料數，p表資料的維度，以data2.npz為例，n=1000, p=2。此時，若要預測一筆新樣本的值，步驟如下：

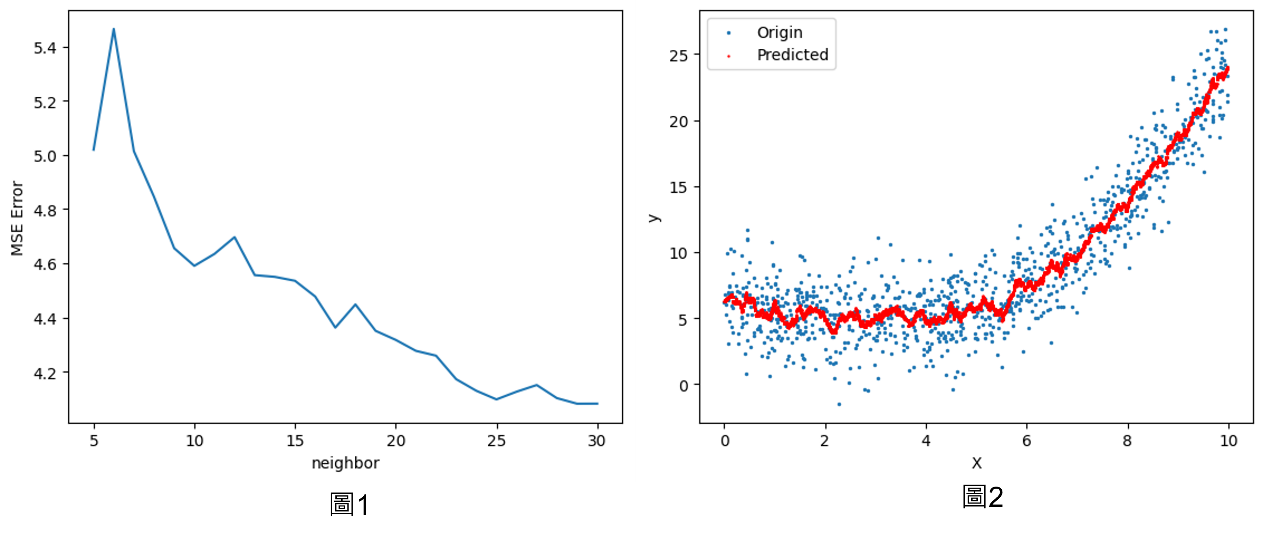
1. 與訓練資料X中的每一筆資料都計算一次歐幾里得距離，並尋找距離最近的K個樣本
2. 使用距離最近的K個樣本進行線性回歸，求解 (這裡的X表K個樣本的x，Y表K個樣本的y)，獲得線性方程式的參數
3. 將樣本代入，獲得

**實驗結果**

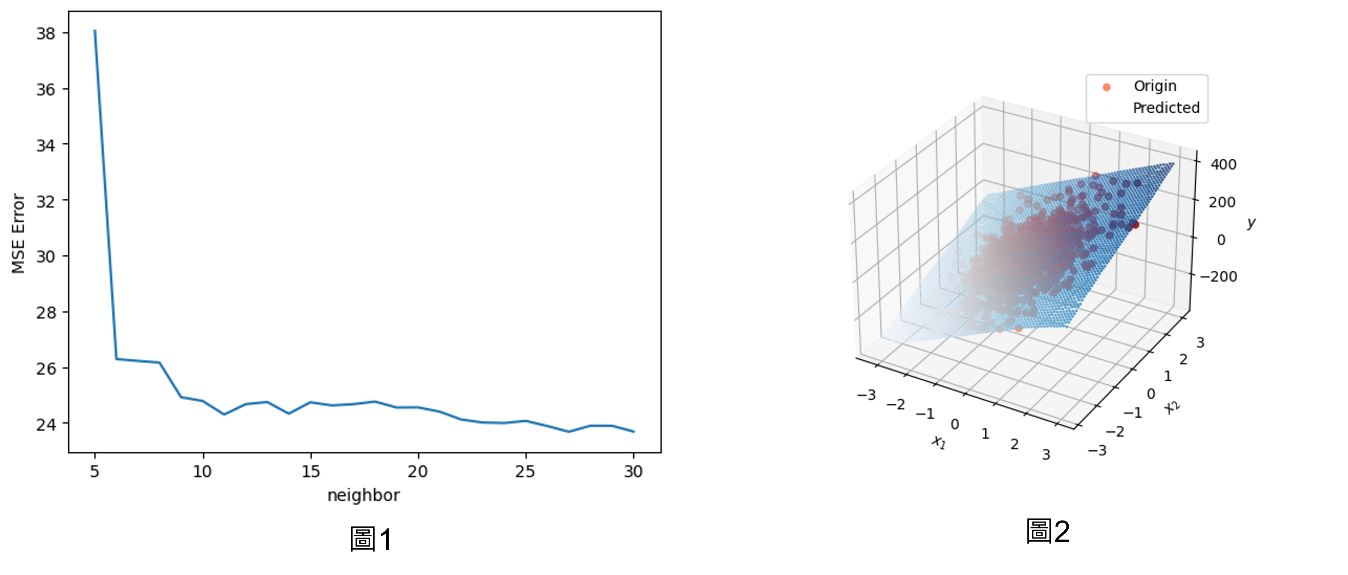
將訓練資料(X, Y)以9:1拆分為訓練集與測試集，並且以KNN linear regression進行測試集的預測，並計算MSE loss。

* 圖1: 設置不同K(neighbors數)時的MSE loss變化，
* 圖2: 根據訓練資料(X, y)，依照X的上下界，在中間均勻取點，產生多筆新資料進行KNN linear regression(K = 20)預測的結果，藍點為原始資料，紅點為新資料

1. Data1.npz:



1. Data2.npz:



**2 Locally weighted regression**

**實現方法**

存在訓練資料(X, Y)，其中X=()，Y)，，n表資料數，p表資料的維度，以data2.npz為例，n=1000, p=2。此時，若要預測一筆新樣本的值，步驟如下：

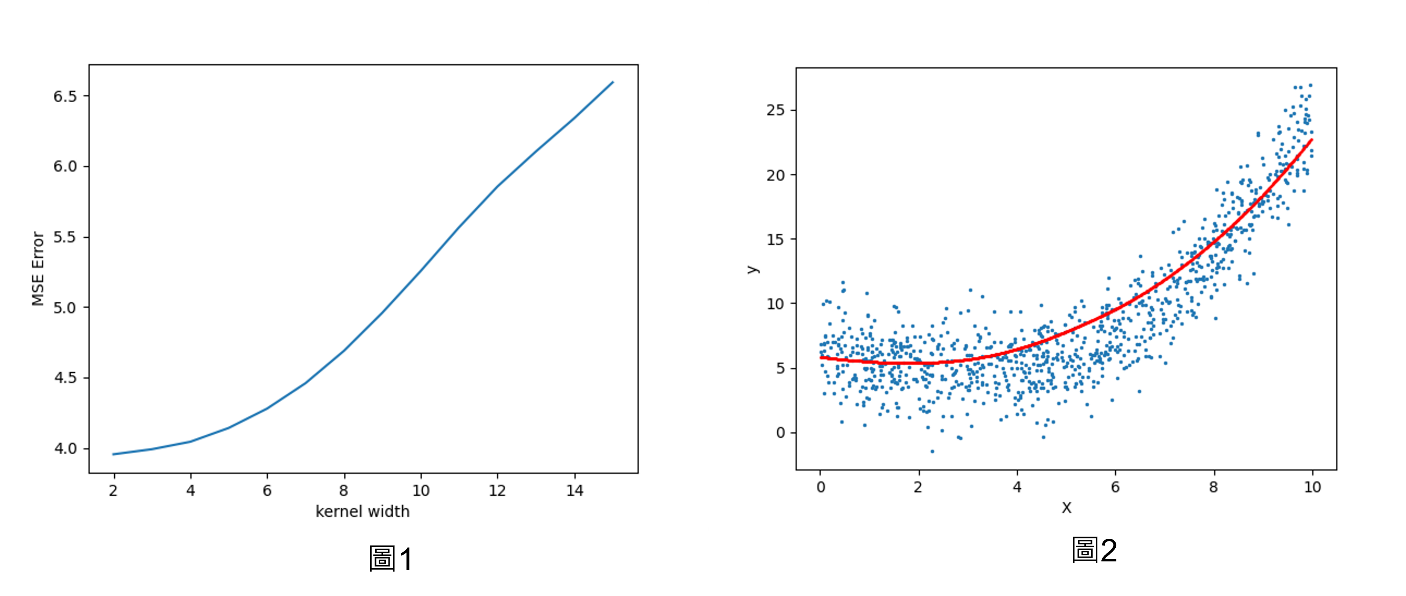
1. 與訓練資料X中的每一筆資料都計算一次歐幾里得距離，將計算出的距離代入quadratic kernel，獲得weight ，k為kernel width
2. 使用該weight w代入方程式(相當於解最佳化式子)，並求解
3. 計算獲得一條線性方程式的權重，
4. 將樣本代入，獲得

**實驗結果**

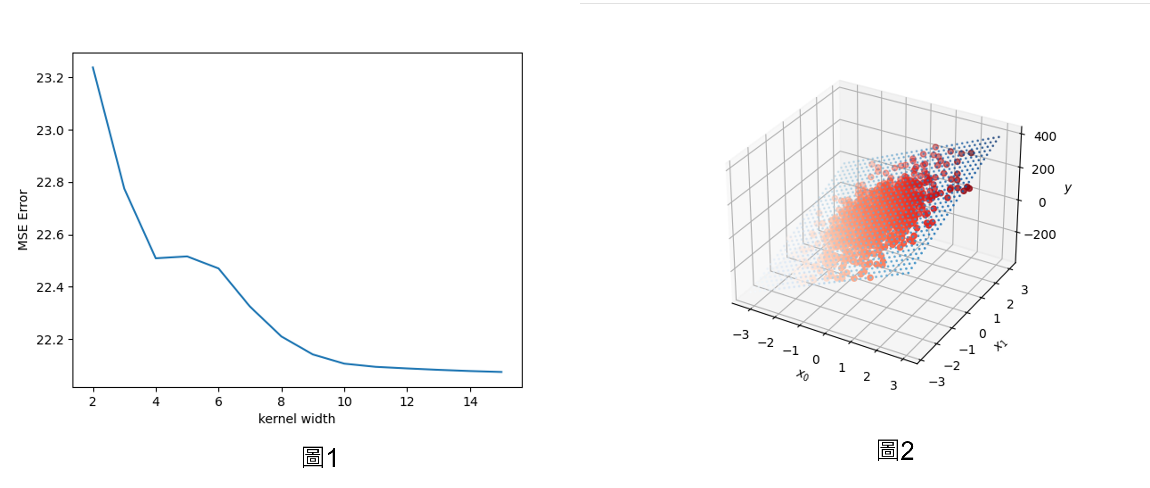
將訓練資料(X, Y)以9:1拆分為訓練集與測試集，並且以Locally weighted regression進行測試集的預測，並計算MSE loss。

* 圖1: 設置不同K(kernel width)時的MSE loss變化，
* 圖2: 根據訓練資料(X, y)，依照X的上下界，在中間均勻取點，產生多筆新資料進行Locally weighted regression (K = 10)預測的結果，藍點為原始資料，紅點為新資料

1. Data1.npz:



1. Data2.npz:



**3 Other Method: K-nearest-neighbors regression**

**實現方法**

存在訓練資料(X, Y)，其中X=()，Y)，，n表資料數，p表資料的維度，以data2.npz為例，n=1000, p=2。此時，若要預測一筆新樣本的值，步驟如下：

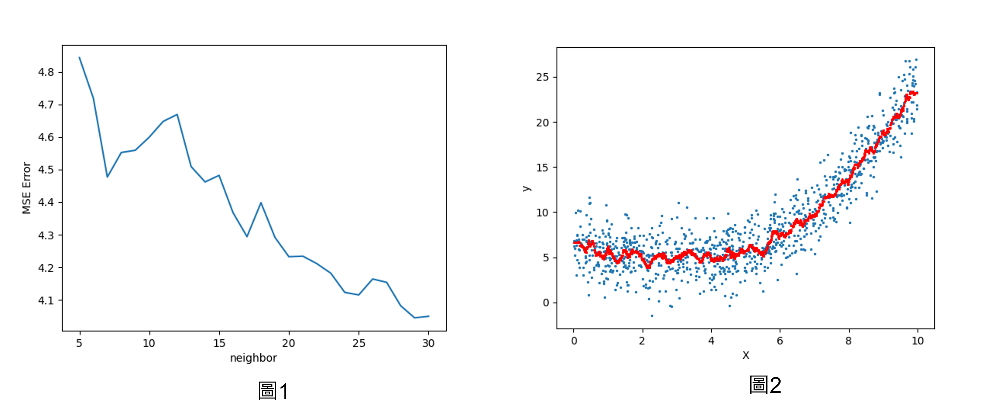
1. 與訓練資料X中的每一筆資料都計算一次歐幾里得距離，並尋找距離最近的K個樣本
2. 使用距離最近的K個樣本之y計算平均 (這裡的K只含距離最近的K個樣本)
3. 此平均值即為樣本對應的

**實驗結果**

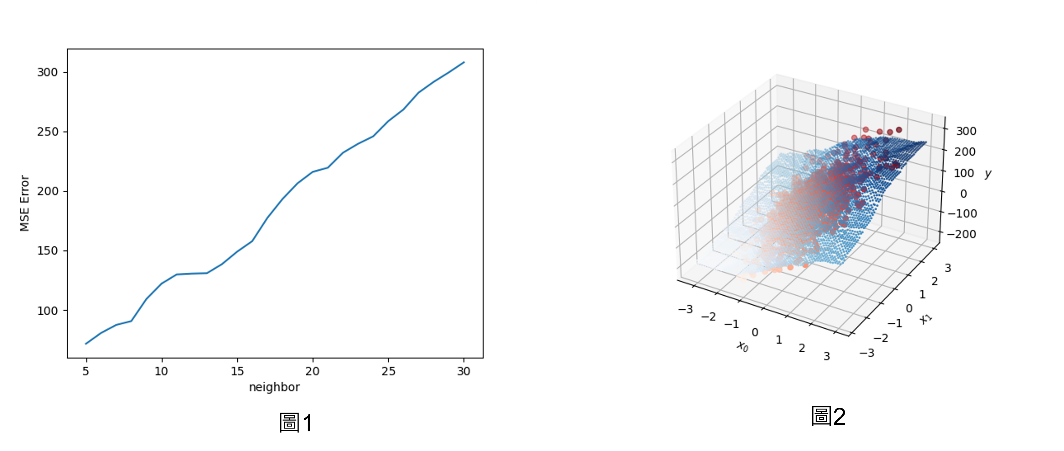
將訓練資料(X, Y)以9:1拆分為訓練集與測試集，並且以KNN regression進行測試集的預測，並計算MSE loss。

* 圖1: 設置不同K(neighbors數)時的MSE loss變化，
* 圖2: 根據訓練資料(X, y)，依照X的上下界，在中間均勻取點，產生多筆新資料進行KNN regression(K = 20)預測的結果，藍點為原始資料，紅點為新資料

1. Data1.npz:



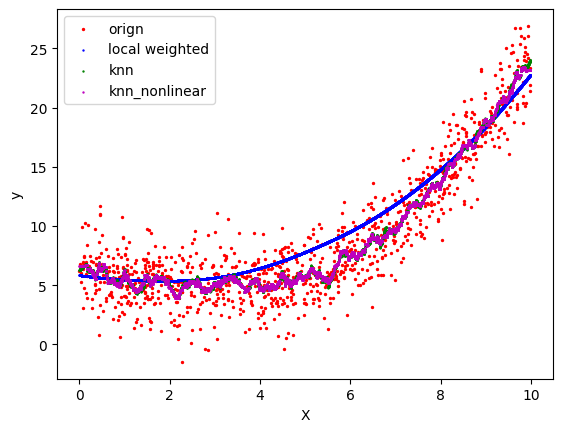
1. Data2.npz:



**4 比較**

* Data1.npz:

下圖是將結果放在同一圖上的結果，可以很明顯地看到KNN k=20(綠點)與KNN nonlinear(紫點)的曲線大致相同，且同為不連續的曲線，而local weighted(藍點)則是連續的曲線

****

* Data2.npz:

下圖是將結果放在同一圖上的結果，在二維的例子中，可以很明顯地發現KNN(綠點)與KNN nonlinear(紫點)的平面不太一致，反而 local weighted(藍點)與KNN綠點)的平面較為接近，這個實驗結果也跟上面loss的結果大致相同，經由觀察上面的MSE loss中，可以發現KNN regression在二維的例子(data2.npz)中MSE Loss很明顯地較KNN linear regression與local weighted regression差上許多

