

雖然KNN是惰性算法，不代表沒有訓練誤差。模型在訓練集上的誤差就是訓練誤差。因此報告將Training Error以及Testing Error一同做討論。

由模型訓練結果圖，可大致分三種來做討論:

1. 前半部(約為K=1~6) **Over-fitting:**

當k=1時，對於訓練集來說，每個點都會被自己本身標上標籤，所以訓練誤差為0。當k>1時，訓練集的孤立點被錯誤分類，因為它們周圍都是其他類別的點。

前半部k取較小時，由於訓練資料少，導致過度符合Training set的資料特性，使得其無法預測較為普遍的資料集，儘管Training Error很低，但是推廣能力不佳，判斷此為Over-fitting問題。

1. K = 9 **Exact the perfect point:**

此點的Testing Error最低，為推廣性最佳的情況。

1. K = 15 ~ 20 **Under-fitting:**

當k越大，Testing Error 和 Training Error 開始明顯示增加，預測結果開始是趨向資料數量最多的那一類(K 過大，忽略太多局部資訊)，等於喪失KNN預測的效果，表示此模型開始發生under-fitting問題。

由此可見，KNN之優點為精度高、對異常值不敏感、無資料輸入假定。

然而缺點便是訓練模型很依賴訓練集資料。