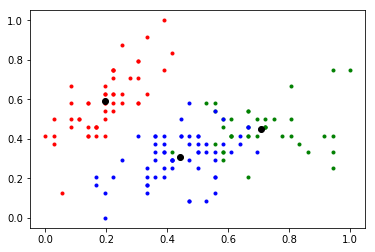
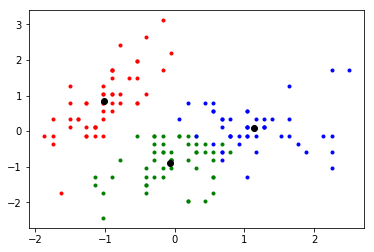
Practice 4

1. Scaling\_Minimum: 29.2788527313(wicd最小值)



將raw data使用scaling方法前處理後，將資料Map到0~1，可以發現有些點會靠近其他群，代表錯誤分群，可以發現將資料映射到0~1，會讓資料差異變得更小，在計算距離時，容易錯判彼此的距離，導致分群錯誤。

1. Standard\_Minimum: 129.67408038(wicd最小值)



運用standard的方式可以發現資料分群較無誇張的錯誤(與上圖比較)

Practice 5

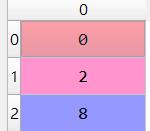
* 前處理1~3

1. 首先產生一個150\*149矩陣T，分別記錄每個點與其他149個點的距離  
   (採用SSD計算距離)。



1. 在來產生150\*K矩陣K\_nearest，記下該點最近的K個index，例如:k=3代表，第一點資料的最近3個點即是17,4,39，再用target[17], target[4], target[39]找出Truth值，在做投票選擇(到step3)。



1. 因此次投票需要紀錄每次選擇的所有值，所以需要建立150個票倉來進行投票，紀錄每次的投票結果，如右圖。
2. 最後將每次的投票結果丟入到該次knn的confusion Metrix中，如下表：

* 表：Confusion Metrix

|  |  |
| --- | --- |
| 1nn | 2nn |
| 3nn | 4nn |
| 5nn | 6nn |
| 7nn | 8nn |
| 9nn | 10nn |

1. 從表中可以觀察出有五個點被錯誤分類，我們拿Truth和Predict值相互比較，來計算error count的總數。發現錯誤數量一致。

