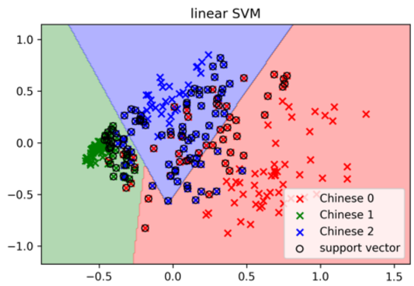
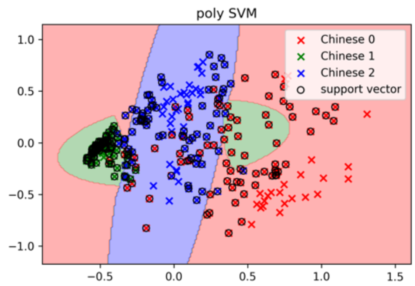
# Machine Learning 2021 Homework 3

310511048 張奕廷

1. **Support Vector Machine (SVM)**
    1. **Discussion.** For one-versus-the-rest decision approach, since each classifier discriminates one class from the others, we only need to build $N$ classifiers for $N$-class classification. On the other hand, one-versus-one approach needs a total of $\frac{N(N-1)}{2}$ classifiers for $N$-class classification because it splits the problem into one binary classification problem for each pair of classes. However, one-versus-the-rest may have larger bias (undefined region) than one-versus-one, while one-versus-one may be more time-consuming since it requires more classifiers. I still choose one-versus-one approach for this problem, because there only exists 3 classes in the dataset, and thus the number of classifiers for both approaches are 3.
    2. Decision boundary and support vectors for SVM with different kernel.




    3. **Discussion.** The decision boundary with polynomial kernel seems to be more flexible, unlike the ones with linear kernel being all straight lines. Theoretically, SVM with polynomial kernel should have better ability to fit the data when the distribution is more complex. However, we can see that SVM with polynomial kernel still misclassifies a considerable amount of samples in this case, especially for class 0 and 2. I observe that in this dataset, samples from class 1 mostly lie in the same region after PCA (Principal Component Analysis), while class 0 and class 2 have quite a few samples mixed together. Hence, I consider the model is somehow confused and cannot give an accurate decision boundary. From this example, we can infer that there is no such classifier which can always have better performance, it all depends on the data distribution.

## 2. Gaussian Mixture Model

1. Table of the estimated $\{\mu_k\}_{k=1}^{K}$ by K-means.

$K = 3$

| K | R | G | B |
|---|---|---|---|
| 0 | 189 | 186 | 177 |
| 1 | 77 | 118 | 106 |
| 2 | 18 | 31 | 35 |

$K = 7$

| K | R | G | B |
|---|---|---|---|
| 0 | 51 | 99 | 82 |
| 1 | 81 | 124 | 112 |
| 2 | 176 | 174 | 163 |
| 3 | 31 | 52 | 59 |
| 4 | 122 | 145 | 138 |
| 5 | 226 | 217 | 208 |
| 6 | 9 | 14 | 18 |

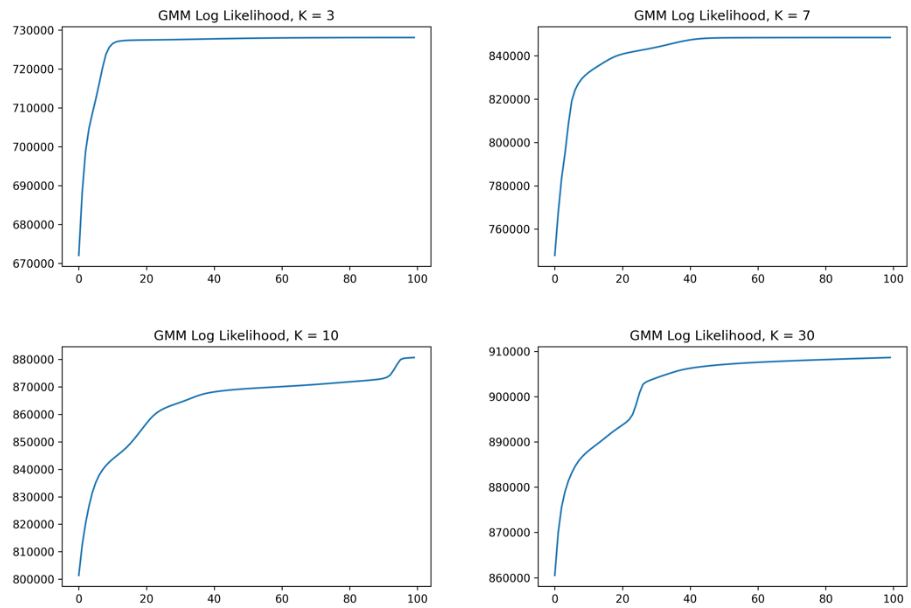
$K = 10$

| K | R | G | B |
|---|---|---|---|
| 0 | 105 | 138 | 125 |
| 1 | 8 | 12 | 17 |
| 2 | 36 | 105 | 72 |
| 3 | 27 | 44 | 52 |
| 4 | 73 | 105 | 164 |
| 5 | 192 | 185 | 174 |
| 6 | 231 | 221 | 213 |
| 7 | 149 | 157 | 147 |
| 8 | 72 | 120 | 100 |
| 9 | 56 | 76 | 84 |

$K = 30$

| K | R | G | B |
|---|---|---|---|
| 0 | 168 | 165 | 149 |
| 1 | 144 | 147 | 131 |
| 2 | 210 | 198 | 185 |
| 3 | 12 | 28 | 61 |
| 4 | 10 | 16 | 22 |
| 5 | 81 | 102 | 95 |
| 6 | 166 | 179 | 196 |
| 7 | 43 | 91 | 70 |
| 8 | 30 | 48 | 92 |
| 9 | 36 | 45 | 48 |
| 10 | 73 | 127 | 104 |
| 11 | 225 | 215 | 207 |
| 12 | 21 | 30 | 33 |
| 13 | 140 | 158 | 157 |
| 14 | 118 | 134 | 127 |
| 15 | 105 | 157 | 139 |
| 16 | 4 | 6 | 8 |
| 17 | 102 | 138 | 189 |
| 18 | 70 | 102 | 162 |
| 19 | 85 | 142 | 121 |
| 20 | 243 | 234 | 226 |
| 21 | 101 | 117 | 110 |
| 22 | 52 | 135 | 97 |
| 23 | 190 | 181 | 164 |
| 24 | 47 | 71 | 125 |
| 25 | 52 | 112 | 84 |
| 26 | 23 | 108 | 68 |
| 27 | 66 | 85 | 81 |
| 28 | 18 | 74 | 49 |
| 29 | 51 | 63 | 65 |

2. Learning curve for log likelihood by Gaussian mixture model (GMM).

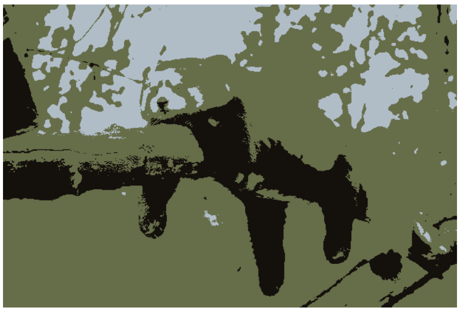



3. Resulting images for $K = 3, 7, 10, 30$.

**<u>K-means</u>**

| $K = 3$ | $K = 7$ |
|---|---|
|  |  |
| $K = 10$ | $K = 30$ |
|  |  |

**<u>Gaussian Mixture Model</u>**

| $K = 3$ | $K = 7$ |
|---|---|
|  |  |
| $K = 10$ | $K = 30$ |
|  |  |

4. **Discussion.** The crucial factor that affects the output image between K-means and Gaussian mixture model (GMM) is the value of $K$, indicating the number of clusters. Firstly, from the results in 2.3, we can see that as $K$ increases, the image by both K-means and Gaussian mixture model (GMM) become more similar to the original one. Also, we can see that when $K$ is small, not only the image looks colorless, but also the output image between two clustering methods looks significantly different. Nevertheless, as $K$ increases, the differences between two output image seems to reduce.