Ömer Alperen Katı
Bogazici University, Electric Electronic Eng., 24.11.2022

EE 573 Pattern Recognition - Project 3 Report

As it is done in the first two projects, both PCA and MDA are applied to the dataset. For PCA, the reduced feature space dimension is selected to be 61.

PART 1: Maximum likelihood estimation (MLE)

After applying MLE, the unknown parameters , μ_c and Σ_c, are determined and given in Appendix I. Using MLE, each test sample is assigned to a class as shown in Fig. 1.

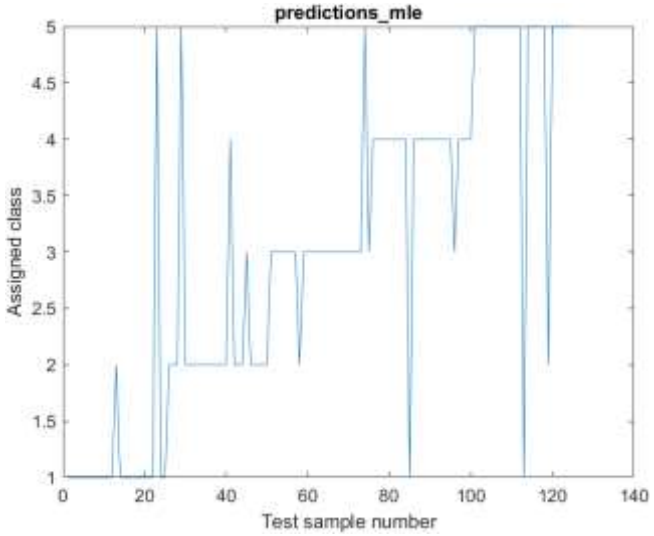

Fig 1: Assigned class of each test sample after MLE (125 in total). Each class has 25 samples and subsets are ordered from w_1 to w_5 , consecutively.

$P(error|x)$ for EM is given for each test sample as shown in Fig. 2.

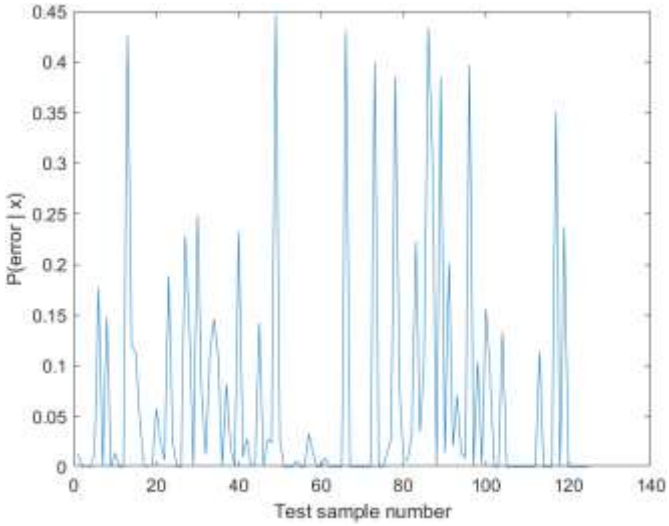

Fig. 2: $P(error \mid x)$ for each test sample after MLE (125 in total). Each class has 25 samples and subsets are ordered from w_1 to w_5 , consecutively.

Taking the mean of the data in Fig. 2. Probability of error is calculated as

$$P(error) \cong \frac{1}{5}\sum_{c=1}^{5}\frac{1}{||T_c||}\sum_{x\in T_c} P(error \mid x) = 0.0674$$

Recall and precision rates for each class are given in Fig. 3 and Fig. 4, consecutively.

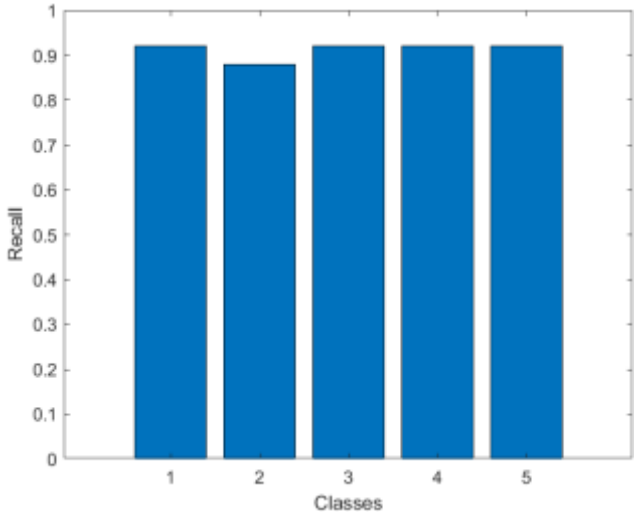

Fig 3: Recall rates for all classes after MLE.

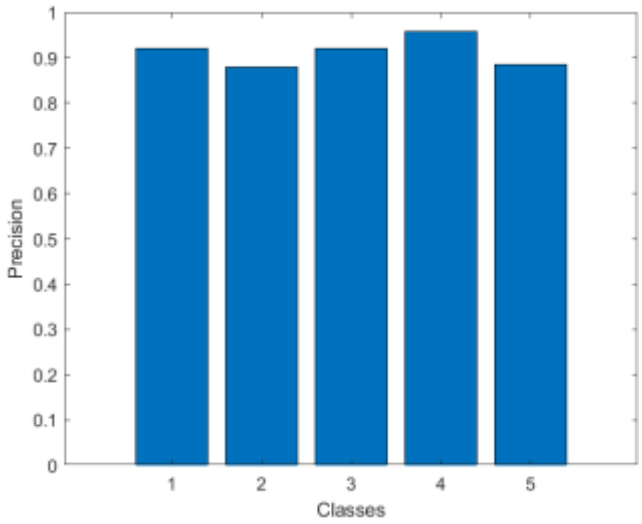

Fig 4: Precision rates for all classes after MLE.

Note that all precision and recall rates are around 0.9. Also, $P(error)$ was found to be 0.0674. Thus, both performance and confidence of MLE algorithm is satisfactory.

PART 2: Expectation maximization (EM)

For the initial values, the results from PART 1 are used instead of that of k-means function. Also, the part that updates the Σ_c is commented out. This because we are required to take Σ_c as constant, determined by the MLE part. To stop the algorithm following criterion is used:

```
sum(abs(mu-old_mu),'all')
```

Using EM, each test sample is assigned to a class as shown in Fig. 5.

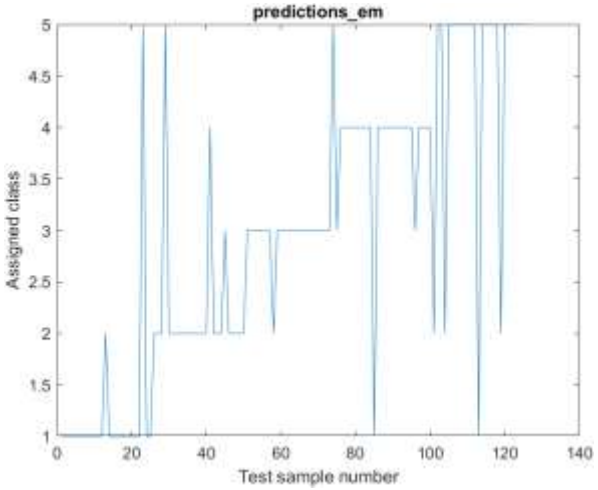

Fig 5: Assigned class of each test sample after EM (125 in total). Each class has 25 samples and subsets are ordered from w_1 to w_5 , consecutively.

$P(error|x)$ for EM is given for each test sample as shown in Fig. 6.

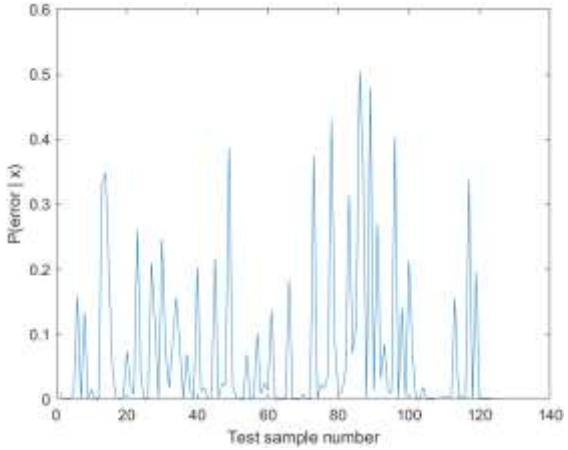

Fig. 6: $P(error \mid x)$ for each test sample after EM (125 in total). Each class has 25 samples and subsets are ordered from w_1 to w_5 , consecutively.

Taking the mean of the data in Fig. 6. Probability of error is calculated as

$$P(error) \cong \frac{1}{5}\sum_{c=1}^{5}\frac{1}{||T_c||}\sum_{x\in T_c} P(error \mid x) = 0.0743$$

Recall and precision rates for each class are given in Fig. 7 and Fig. 8, consecutively.

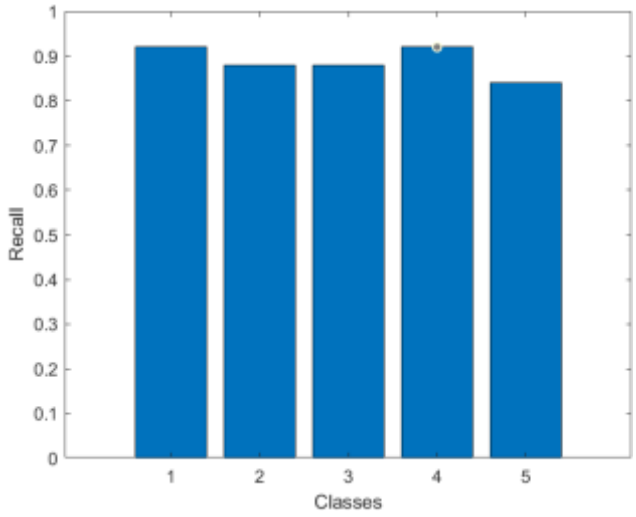

Fig 7: Recall rates for all classes after EM.

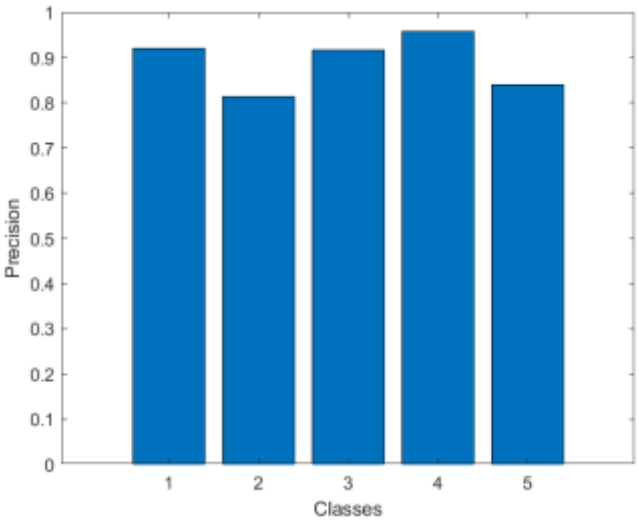

Fig 8: Precision rates for all classes after EM.

Note that all precision and recall rates are around 0.85. Also, $P(error)$ was found to be 0.0743. Thus, both performance and confidence of EM algorithm is satisfactory. But compared to MLE, EM performed slightly worse.

Decision rule for MLE

$$P(x|w_j) = N(\mu_{jm}, \Sigma_j) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_{jm})^T \Sigma_j^{-1}(x-\mu_{jm})\right)$$

$$\arg\max_j \ln P(w_j) p(x_i|w_j)$$

$$\arg\max_j \left[\ln P(w_j) + \ln\left(-\frac{1}{2}(x-\mu_{jm})^T \Sigma^{-1}(x-\mu_{jm})\right) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_{jm}|\right]$$

$$\boldsymbol{\arg\max_j \left[\ln\left(-\frac{1}{2}(x-\mu_{jm})^T \Sigma^{-1}(x-\mu_{jm})\right) + \frac{1}{2}\ln|\Sigma_{jm}|\right]}$$

Decision rule for EM

$$P(x|w_j) = \sum_{m=1}^{2} \gamma_{jm} N(\mu_{jm}, \Sigma_j) = \sum_{m=1}^{2} \gamma_{jm} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_{jm})^T \Sigma_j^{-1}(x-\mu_{jm})\right)$$

$$\arg\max_j P(w_j) P(x_i|w_j)$$

$$\boldsymbol{\arg\max_j P(w_j) \sum_{m=1}^{2} \gamma_{jm} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{jm}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_{jm})^T \Sigma^{-1}(x-\mu_{jm})\right)}$$

Appendix I

```
MATLAB Command Window                                          Page 1


>> param_struct_mle.mu
ans =
   2.0232e-02  -7.2323e-03   8.4883e-04  -4.0643e-03
ans =
   1.8973e-02  -7.9864e-03  -4.4582e-04  -5.4500e-03
ans =
   1.6781e-02  -9.9550e-03  -2.8135e-04  -3.9675e-03
ans =
   1.8137e-02  -6.0638e-03  -1.0950e-03  -3.9223e-03
ans =
   2.2349e-02  -9.1095e-03  -9.5367e-04  -4.0766e-03
>> param_struct_mle.sigma
ans =
    4.2398e-07  -1.3937e-07   8.3294e-08   1.7742e-07
   -1.3937e-07   2.5769e-07   4.5364e-08  -9.2402e-08
    8.3294e-08   4.5364e-08   3.2959e-07   5.9618e-08
    1.7742e-07  -9.2402e-08   5.9618e-08   2.6744e-07
ans =
    6.3108e-07  -9.5968e-08   2.2072e-08  -5.6482e-08
   -9.5968e-08   4.0936e-07  -1.4797e-08   1.8372e-07
    2.2072e-08  -1.4797e-08   1.5560e-07   5.1117e-08
   -5.6482e-08   1.8372e-07   5.1117e-08   4.8586e-07
ans =
    5.3519e-07   4.6942e-07  -1.0973e-08  -1.6430e-07
    4.6942e-07   7.1409e-07  -3.5546e-08  -1.6791e-07
   -1.0973e-08  -3.5546e-08   8.1045e-08   2.9020e-08
   -1.6430e-07  -1.6791e-07   2.9020e-08   2.9044e-07
ans =
    3.4754e-07  -1.0665e-07   1.2172e-07  -1.5930e-07
   -1.0665e-07   4.9368e-07  -1.5120e-07   1.7262e-07
    1.2172e-07  -1.5120e-07   1.7028e-07  -1.0951e-07
   -1.5930e-07   1.7262e-07  -1.0951e-07   2.7899e-07
ans =
    8.0974e-07  -1.2743e-07  -2.1612e-07   2.0267e-07
   -1.2743e-07   3.0854e-07   1.5618e-07  -9.6029e-08
   -2.1612e-07   1.5618e-07   2.8503e-07  -3.0248e-08
    2.0267e-07  -9.6029e-08  -3.0248e-08   3.1183e-07
>>
```

References

classify_paramteric.m, EM.m, k_means.m, ML.m, MultipleDiscriminantAnalysis.m, PCA.m, plot_process.m functions in DHS toolbox are used in this project.