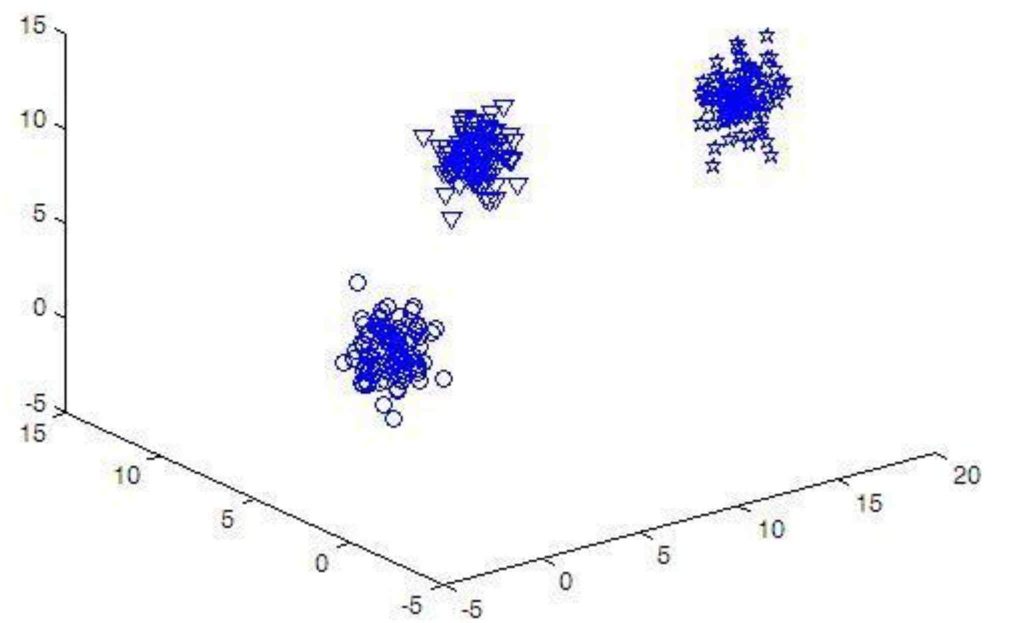
CSE 474
Pattern Recognition Sessional

Lab# 1: Implementation of a Bayesian Classifier

Bayes Classifier

- Assume following training set is given
  - 3 classes
  - 3 features



Bayes Classifier

- Assume following training set is given
  - All numerical data
  - NO assumption for conditionally independence

F1	F2	F2	Class
9.4512	7.3199	6.4664	1
10.7276	9.6067	5.9398	3
10.1960	9.3145	8.3873	1
15.7777	1.5879	11.4440	2
15.8685	2.7902	11.2532	3
14.9448	0.7798	12.7481	2

Bayes Classifier

- Given an unknown sample,

  F = [F1, F2, F3] = [10.1960 9.3145 8.3873]

Predict its class!

Naïve Bayes Classifier

- Given an unknown sample,

  F = [F1, F2, F3] = [10.1960 9.3145 8.3873]

Predict its class!

Calculate posterior probabilities for each class C_i

$P(C_i | F1,F2,F3) \sim = P(C_i)\ P([F1,F2,F3]|C_i)$

Bayes Classifier

- Given an unknown sample,

  F = [F1, F2, F3] = [10.1960 9.3145 8.3873]

Predict its class!

Calculate posterior probabilities for each class C_i

$P(C_i| [F1,F2,F3]) \sim = P(C_i)\, P([F1, F2, F3]|C_i)$

Assign [F1, F2, F3] to C_i if $P(C_i| [F1,F2,F3]) > P(C_j| [F1,F2,F3])$ for all j

Bayes Classifier

- To estimate P([F1, F2, F3]|C_i), assume **F**=[F1, F2, F3] follows multivariate Gaussian distribution

$$P(\mathbf{F} \mid C_i) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_i\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{F}-\boldsymbol{\mu}_i)^t \Sigma_i^{-1}(\mathbf{F}-\boldsymbol{\mu}_i)\right]$$

F1	F2	F2	Class
9.4512	7.3199	6.4664	1
10.7276	9.6067	5.9398	3
10.1960	9.3145	8.3873	1
15.7777	1.5879	11.4440	2
15.8685	2.7902	11.2532	3
14.9448	0.7798	12.7481	2

Bayes Classifier

- Separate training data according to class
- For each class, estimate $\boldsymbol{\mu}_i = [\mu_{i1}, \mu_{i2}, \mu_{i3}]$and covariance matrix $\boldsymbol{\Sigma}_i$ from training samples only

F1	F2	F2	Class
9.4512	7.3199	6.4664	1
10.1960	9.3145	8.3873	1

$\uparrow$ μ_{11} $\quad$ $\uparrow$ μ_{12} $\quad$ $\uparrow$ μ_{13}

- For covariance matrix $\boldsymbol{\Sigma}_i$, you can use library function

Bayes Classifier

- Priors, $P(C_i)$, can be easily estimated by counting the no. of training samples of the corresponding classes

F1	F2	F2	Class
9.4512	7.3199	6.4664	1
10.7276	9.6067	5.9398	3
10.1960	9.3145	8.3873	1
15.7777	1.5879	11.4440	2
15.8685	2.7902	11.2532	3
14.9448	0.7798	12.7481	2

Bayes Classifier

- Training File: train.txt
  - Use this file for training (estimating the Gaussian parameters and priors)
  - File format:
    - first line contains 3 integers m n p. m= no. of features, n= no. of classes, p=no. of samples
    - Each of next p lines contains m feature values followed by its class in integer

Bayes Classifier

- Testing file: test.txt
  - Format is identical but the first line is removed

F1	F2	F2	Class
9.4512	7.3199	6.4664	1
10.7276	9.6067	5.9398	3
10.1960	9.3145	8.3873	1
15.7777	1.5879	11.4440	2
15.8685	2.7902	11.2532	3
14.9448	0.7798	12.7481	2

  - For each vector **F** =[F1, F2, F3] = [9.4512   7.3199 6.4664], find P(**F**|C$_i$) using

$$P(\mathbf{F}\,|\,C_i) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_i\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{F}-\boldsymbol{\mu}_i)^t\,\Sigma_i^{-1}(\mathbf{F}-\boldsymbol{\mu}_i)\right]$$

Bayes Classifier

- Testing file: test.txt
  - Use Bayesian rule to find predicted class

    Assign **F** to C_i if $P(C_i| \mathbf{F}) > P(C_j| \mathbf{F})$ for all j

  - Compare predicted and actual class and calculate % of accuracy

Bayes Classifier

- Output:
  - Identify all misclassified samples and report as follows

    sample no. feature values actual class estimated class

  - % of accuracy