Matlab 10: Quadratic Classifiers

Cheng-Hsin Hsu

National Tsing Hua University

Department of Computer Science

Slides are based on the materials from Prof. Roger Jang

Recap: PDF Modeling

- Goal:
  - Find a PDF (probability density function) that can best describe a given dataset
- Steps:
  - Select a class of parameterized PDF
  - Identify the parameters via MLE (maximum likelihood estimate) based on a given set of sample data
- Commonly used PDFs:
  - Multi-dimensional Gaussian PDF
  - Gaussian mixture models (GMM)

PDF Modeling for Classification

- Procedure for classification based on PDF
  - Training stage: PDF modeling of each class based on the training dataset
  - Test stage: For each entry in the test dataset, pick the class with the max. PDF
- Commonly used classifiers:
  - Quadratic classifiers, with n-dim. Gaussian PDF
  - Gaussian-mixture-model classifier, with GMM PDF

1D Gaussian PDF Modeling

- 1D Gaussian PDF:

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- MLE of μ (mean) and σ^2 (variance)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$$

D-dim. Gaussian PDF Modeling

- D-dim Gaussian PDF:

$$\boxed{g(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}\left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}}$$

- MLE of μ (mean) and Σ (covariance matrix)

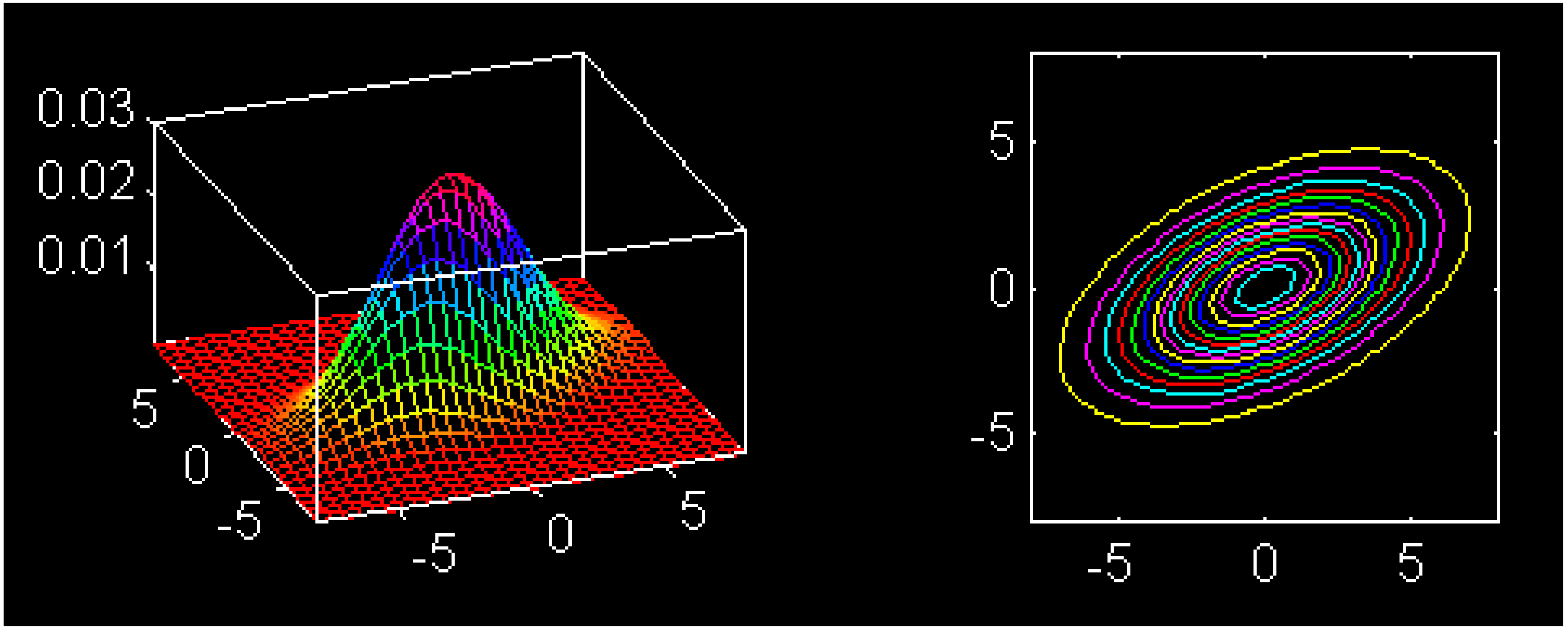
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \mu\right)\left(x_i - \mu\right)^T$$

2D Gaussian PDF

- Bivariate normal distribution:

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0.5 \\ 0.5 & 2 \end{bmatrix}$$



Training and Testing Stages of Quadratic Classifier (QC)

- Training stage
  - Identify the Gaussian PDF of each class via MLE
- Testing stage
  - Assign a sample point to the class C by taking class prior into consideration:

$$\boxed{\widehat{C} = \arg\max_{C} \Pr(C) * Pdf_C(\mathbf{x})}$$

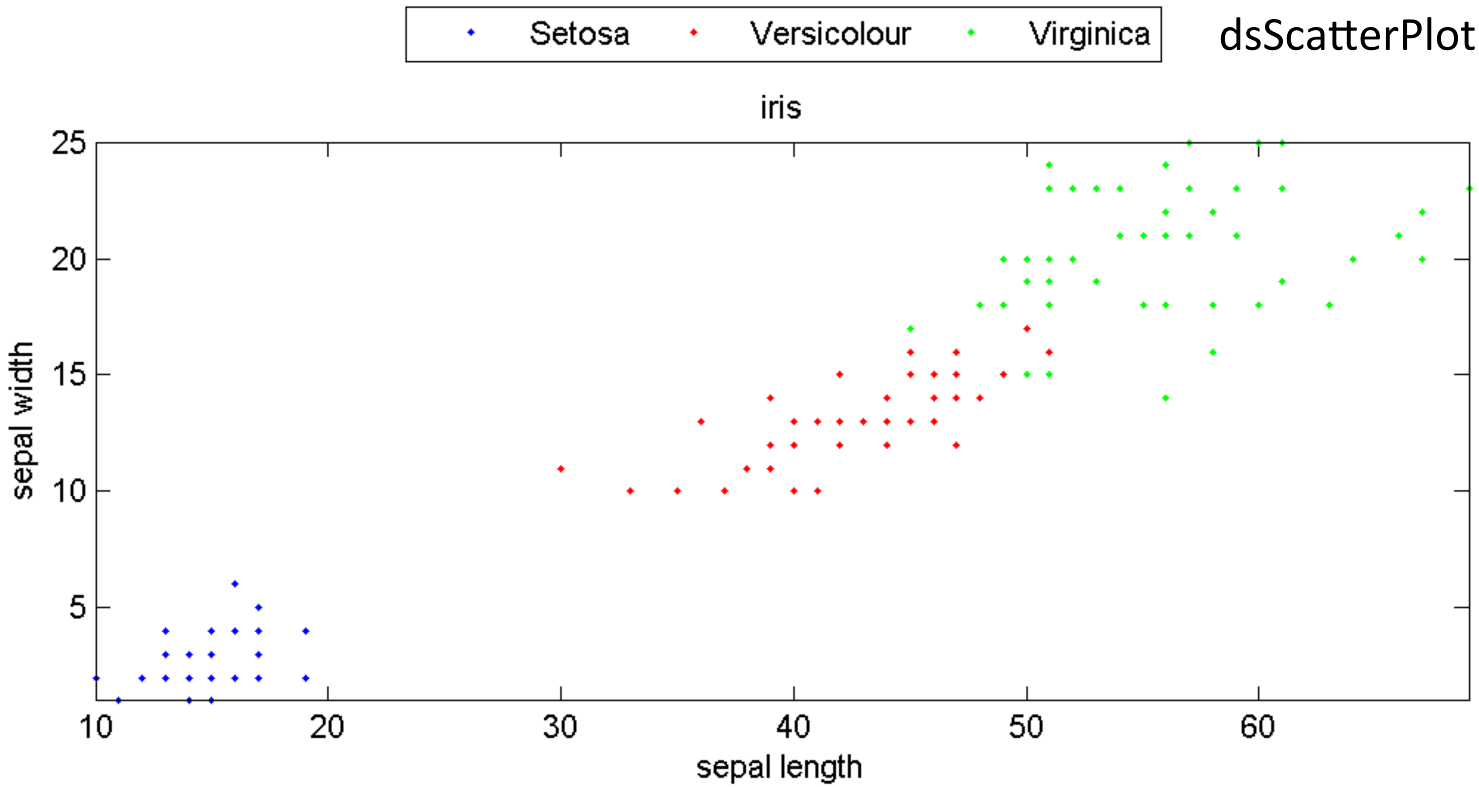
Characteristics of QC

- If each class is modeled by an Gaussian PDF, the decision boundary between any two classes is a quadratic function.
- That is why it is called quadratic classifier.

QC on Iris Dataset (1/3)

- Scatter plot of Iris dataset (with only the last two dim.)

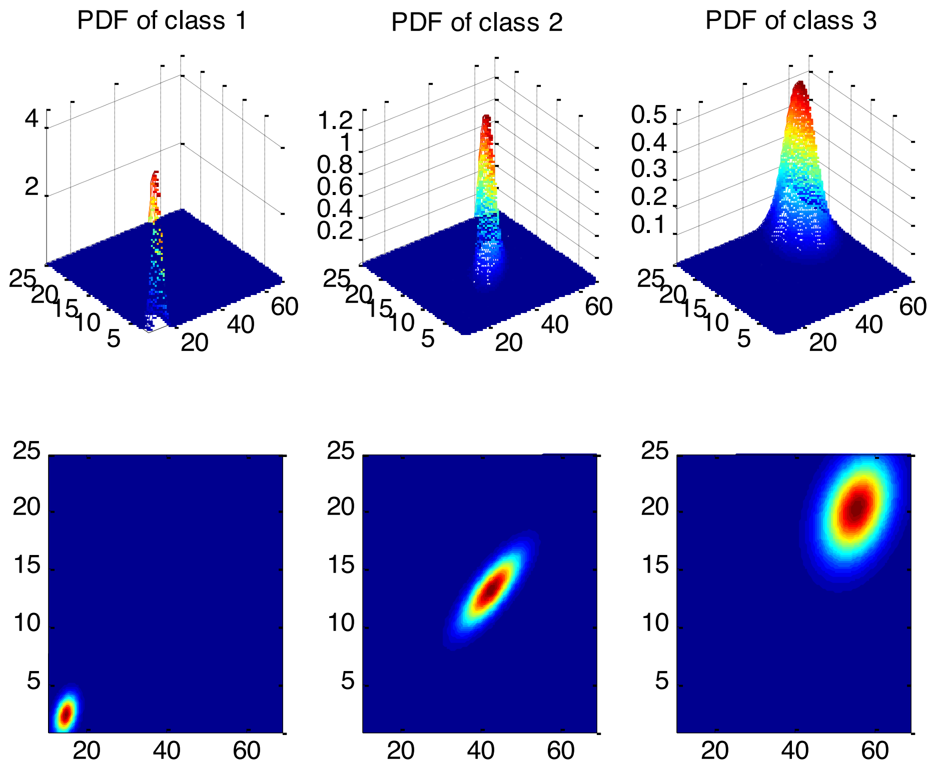
```
DS = prData('iris');
DS.input=DS.input(3:4, :);
dsScatterPlot(DS);
```



QC on Iris Dataset (2/3)

- PDF for each class:

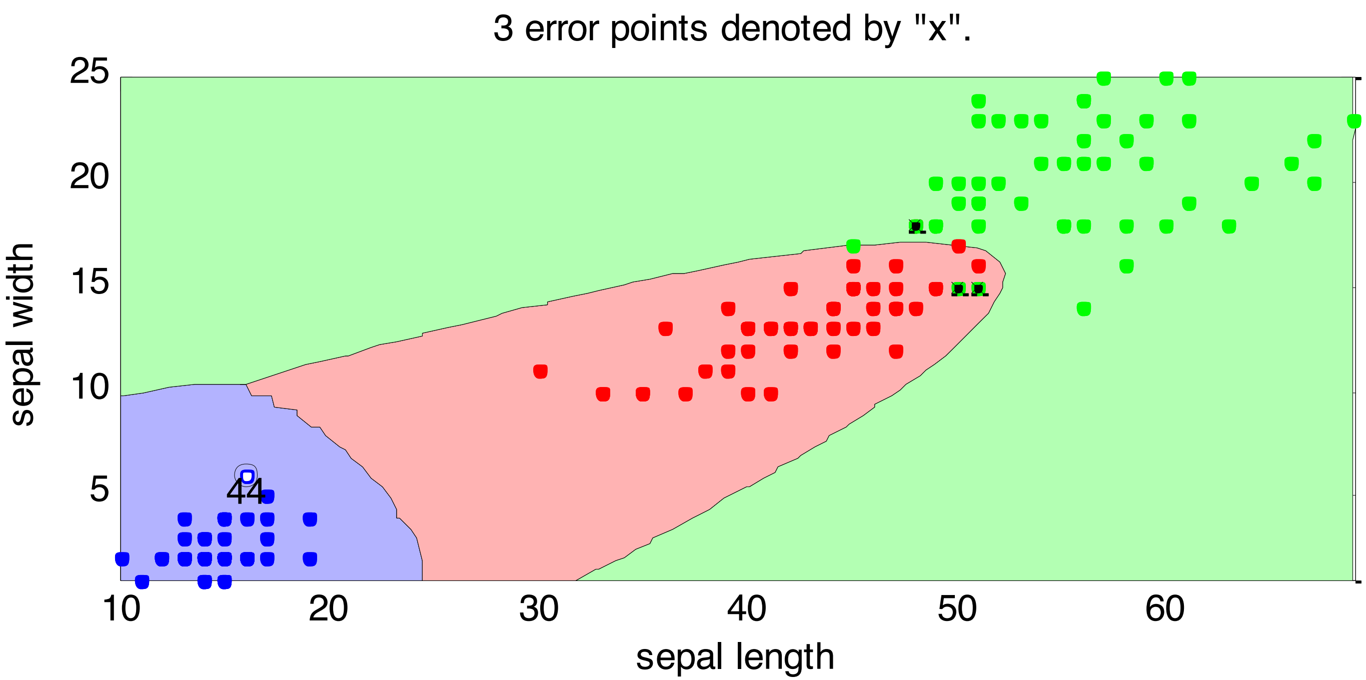
```
DS=prData('iris');
DS.input=DS.input(3:4, :);
[qcPrm, logProb, recogRate, hitIndex]=qcTrain(DS);
qcPlot(DS, qcPrm, '2dPdf')
```



QC on Iris Dataset (3/3)

- Decision boundaries among classes:

```
DS=prData('iris');
DS.input=DS.input(3:4, :);
[qcPrm, logProb, recogRate, hitIndex]=qcTrain(DS);
DS.hitIndex=hitIndex;
qcPlot(DS, qcPrm, 'decBoundary');
```



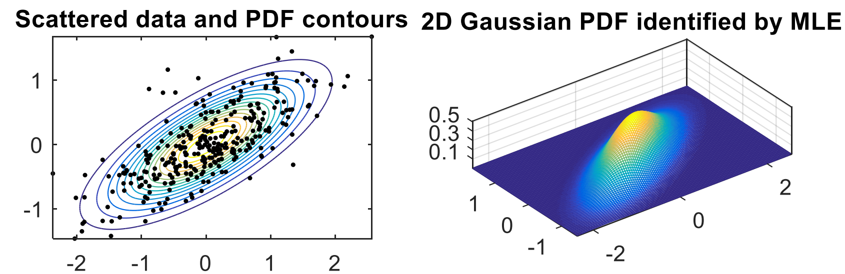
Strength and Weakness of QC

- Strength
  - Efficient computations when the dimension d is small
  - Efficient way to compute leave-one-out cross validation ← what is validation?
- Weakness
  - The covariance matrix (d by d) is big when the dimension d is median large
  - The inverse of the covariance matrix may not exist
  - Cannot handle bi-modal data

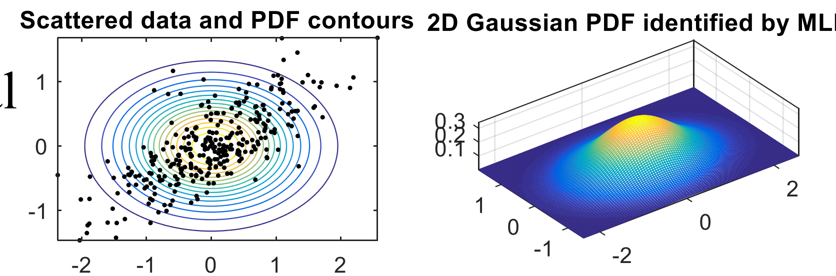
What If No Inverse Matrix (or 0 Determinant)

- How to modify QC such that it won’t deal with a big covariance matrix?
  - Make the covariance matrix diagonal ➔ Equivalent to naïve Bayes classifiers
  - Make the covariance matrix a constant times an identity matrix

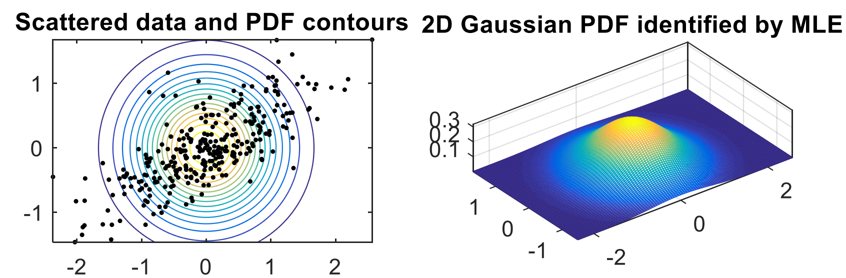
$\Sigma : \text{full}$



$\Sigma : \text{diagonal}$



$\Sigma = \sigma^2 I$



Matlab #9 Homework (M9) 2%

- (1%) Use prData.m to obtain the ABALONE dataset, in which the desired output is the age for abalones. Consider two arbitrary dimensions, and use the Quadratic classifier to obtain the decision boundaries. Please turn in the figure and your observations.
- (1%) **Quadratic classifier**:
  - Why the classifier is named "quadratic"?
  - How do you train a quadratic classifier?
  - How do you evaluate (test) a quadratic classifier?
  - What is the major strength of a quadratic classifier?
  - What is the major weakness of a quadratic classifier?

Questions?