

# Matlab 11: Naive Bayes Classifiers



**Cheng-Hsin Hsu**

*National Tsing Hua University*

*Department of Computer Science*

Slides are based on the materials from Prof. Roger Jang

# Assumptions & Characteristics

- Assumptions:
  - Statistical independency between features
  - Statistical independency between samples
  - Each feature governed by a feature-wise parameterized PDF (usually a 1D Gaussian)
- Characteristics
  - Simple and easy (That's why it's named "naive".)
  - Highly successful in real-world applications regardless of the strong assumptions

# Training and Test Stages of NBC

- Training stage

- Identify class PDF, as follows.

- Identify feature PDF by [MLE for 1D Gaussians](#)
    - Class PDF is the product of all the corresponding feature PDFs

$$Pdf_C(\mathbf{x}) = Pdf_C(x_1) * Pdf_C(x_2) * \dots * Pdf_C(x_d)$$

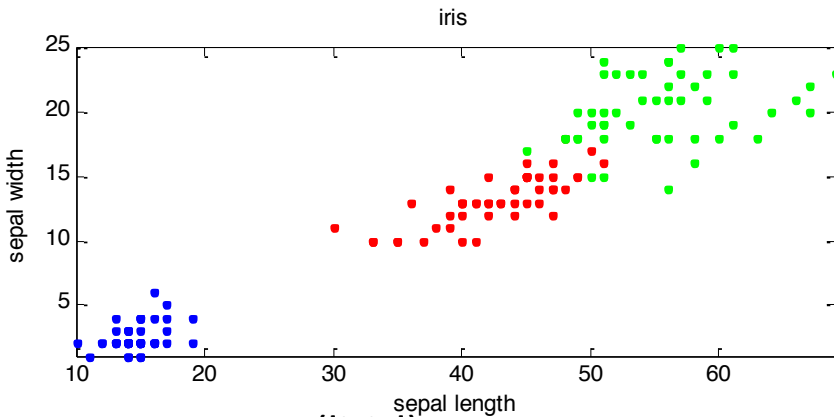
- Test stage

- Assign a sample to the class by taking class prior into consideration:

$$C = \arg \max_C Pr(C) * Pdf_C(\mathbf{x})$$

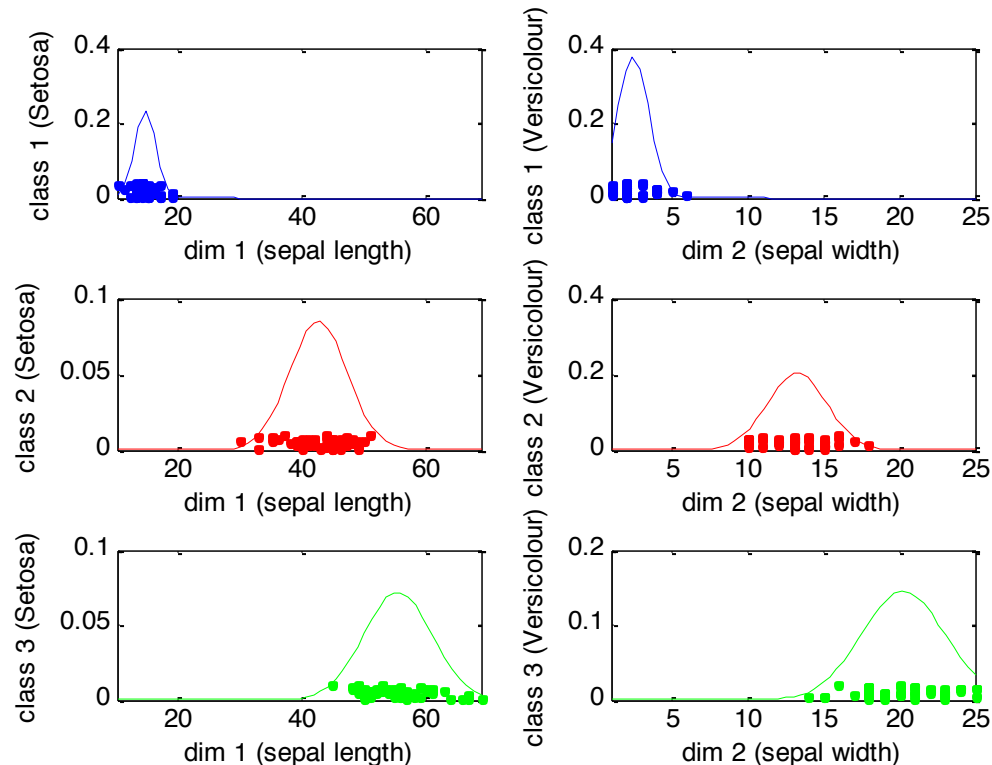
# NBC for Iris Dataset (1/3)

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



```
DS=prData('iris');  
DS.input=DS.input(3:4, :);  
dsScatterPlot(DS);  
figure;  
[nbcPrm, logLike, recogRate,  
hitIndex]=nbcTrain(DS);  
nbcPlot(DS, nbcPrm, '1dPdf');
```

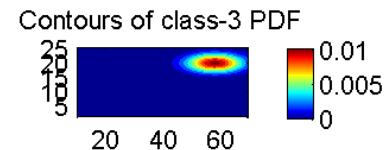
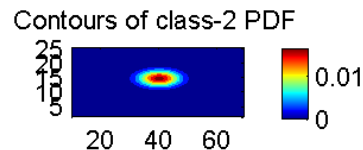
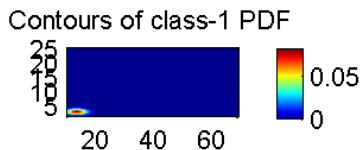
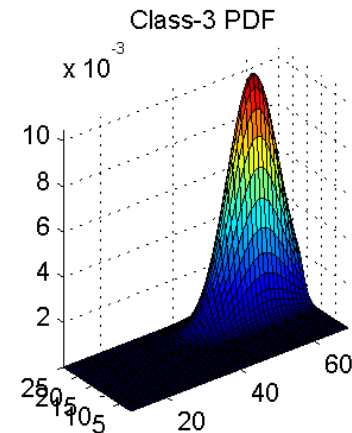
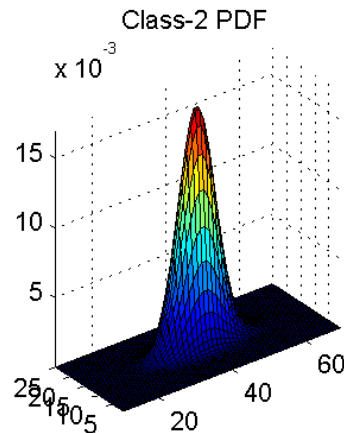
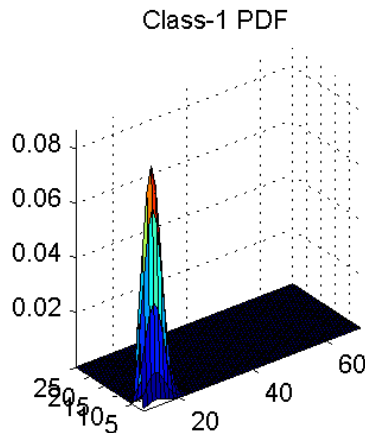
- PDF on each features and each class



# NBC for Iris Dataset (2/3)

- PDF for each class

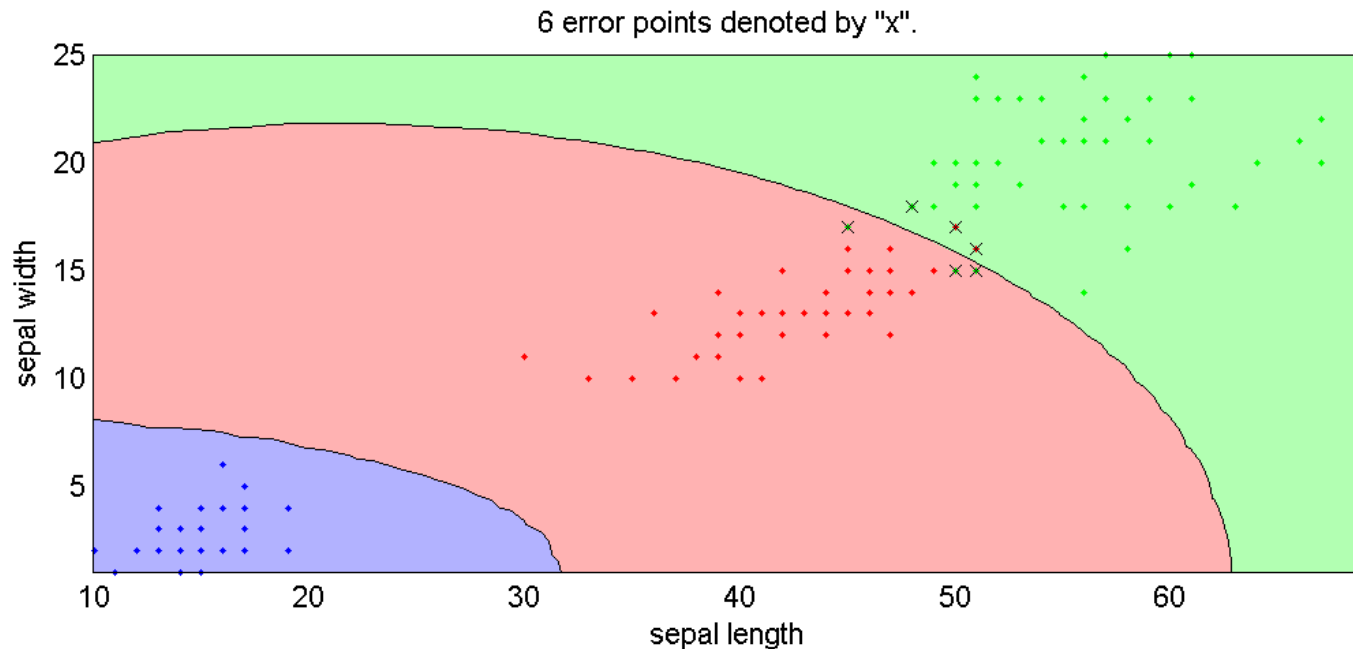
```
DS=prData('iris');  
DS.input=DS.input(3:4, :);  
[nbcPrm, logLike, recogRate, hitIndex]=nbcTrain(DS);  
nbcPlot(DS, nbcPrm, '2dPdf');
```



# NBC for Iris Dataset (3/3)

- Decision boundaries (using the last 2 features)

```
DS=prData('iris');  
DS.input=DS.input(3:4, :);  
[nbcPrm, logLike, recogRate, hitIndex]=nbcTrain(DS);  
DS.hitIndex=hitIndex;  
nbcPlot(DS, nbcPrm, 'decBoundary');
```



# Strength and Weakness of NBC

- Strength
  - Fast computations during training and evaluation
  - Robust than QC
- Weakness
  - Cannot deal with bi-modal data
  - Class boundary not as complex as QC

# Matlab #10 Homework (M10) 2%

---

- (1%) Use `prData.m` to obtain the WINE dataset, in which the desired output is the origin of the WINE (3 classes in total). Consider two arbitrary dimensions, and use the Naive Bayes classifier to obtain the decision boundaries. Please turn in the figure and your observations.
- (1%) **Naive Bayes classifier:**
  - How do you train a naive Bayes classifier?
  - How do you evaluate (test) a naive Bayes classifier?
  - What is the major strength of a naive Bayes classifier?
  - What is the major weakness of a naive Bayes classifier?



# Questions?

---

