K-Nearest Neighbor Classifiers (KNNC)

J.-S. Roger Jang (張智星)

jang@mirlab.org

http://mirlab.org/jang

MIR Lab, CSIE Dept.
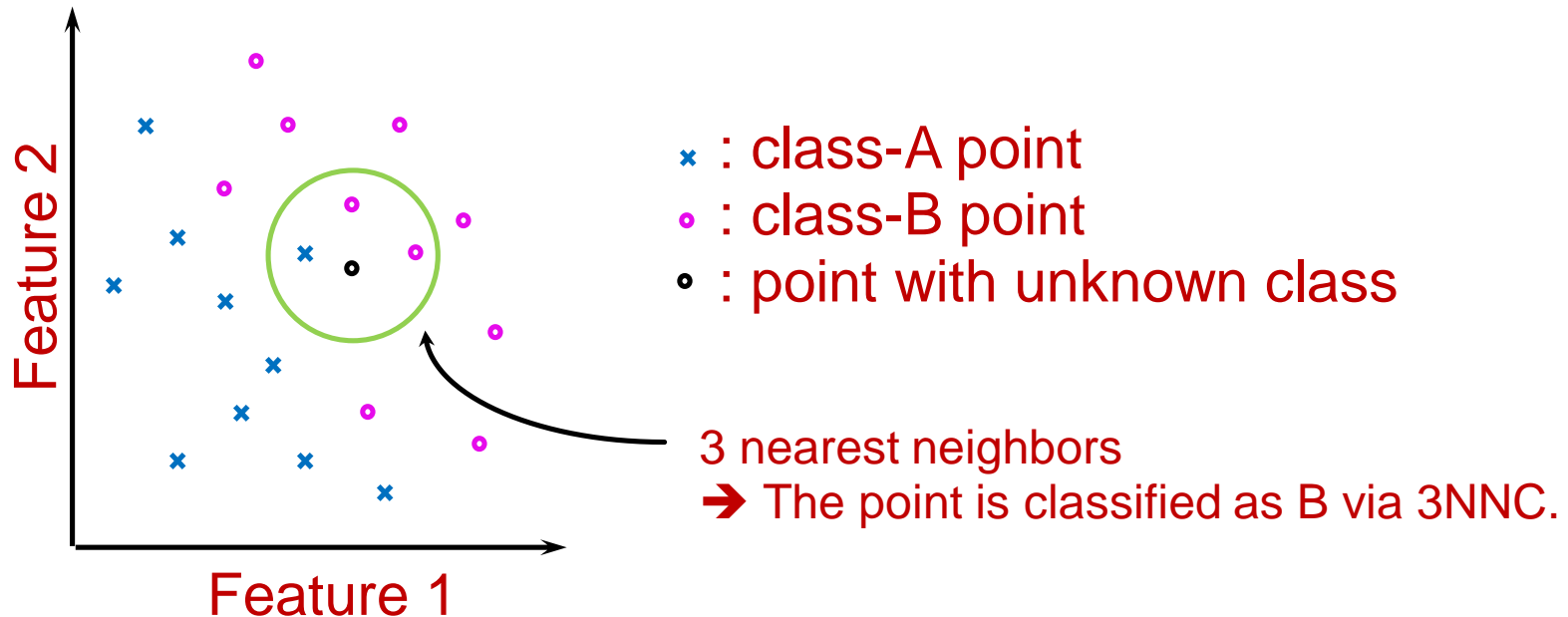
National Taiwan University

2023/9/20
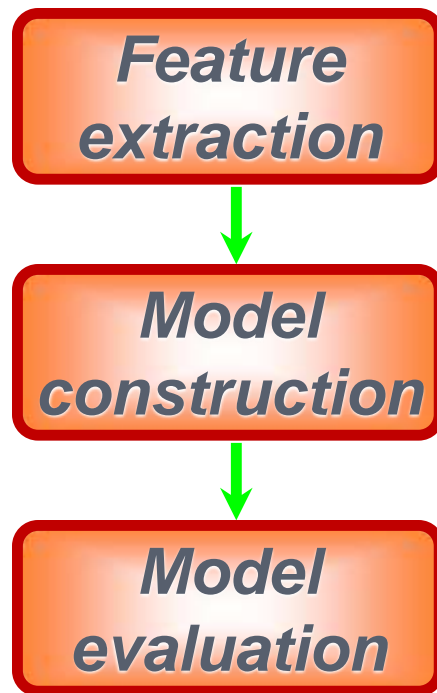
Concept of KNNC

- Concept: 近朱者赤、近墨者黑
- Two Steps: Quiz!
  - Find the first k nearest neighbors of a given point.
  - Determine the class of the given point by voting among k nearest neighbors.

× : class-A point

○ : class-B point

○ : point with unknown class

3 nearest neighbors
➔ The point is classified as B via 3NNC.

Flowchart for KNNC

Flowchart of classification: **KNNC:**
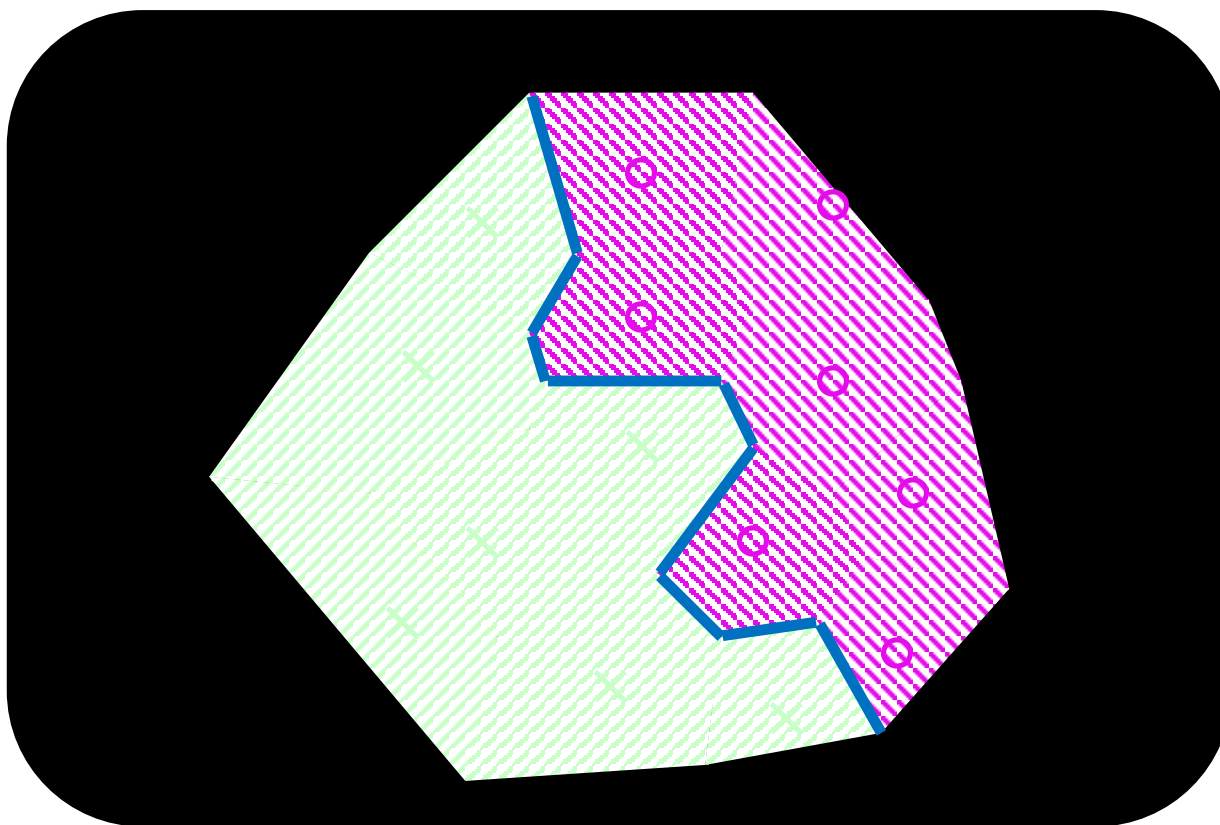

From raw data to features

Clustering (optional)

**KNNC evaluation
on test dataset**

Decision Boundary for 1NNC
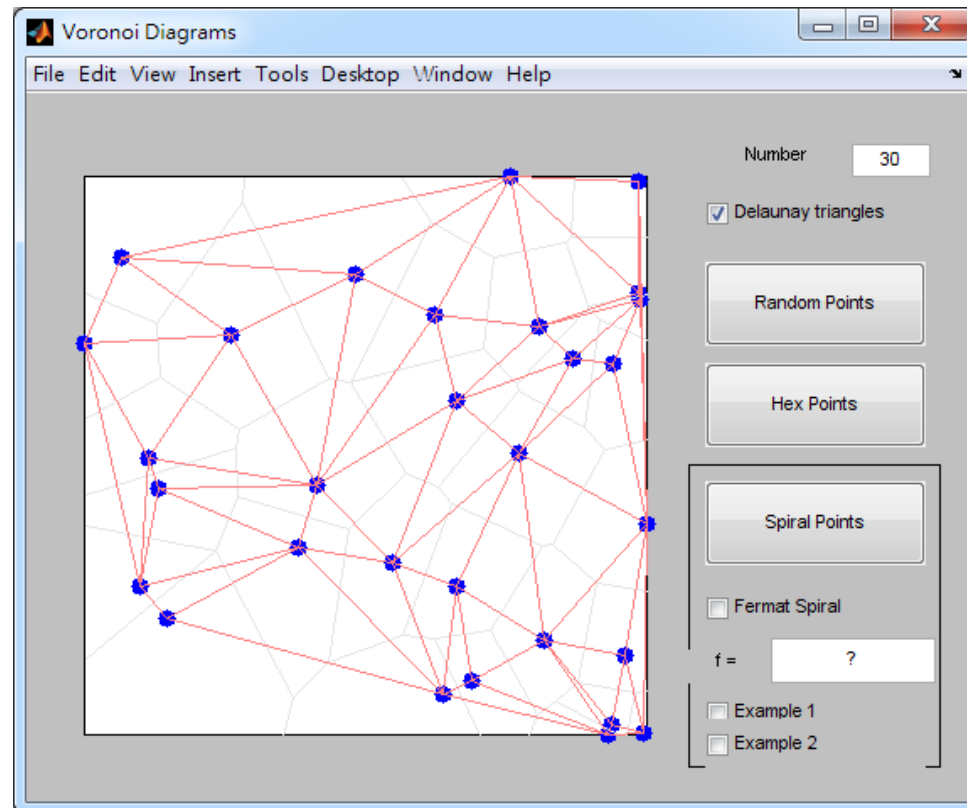
○ Voronoi diagram: piecewise linear boundary

Quiz!

More about Voronoi diagrams

Demos by Cleve Moler

- Cleve's Demos of Delaunay triangles and Voronoi diagram
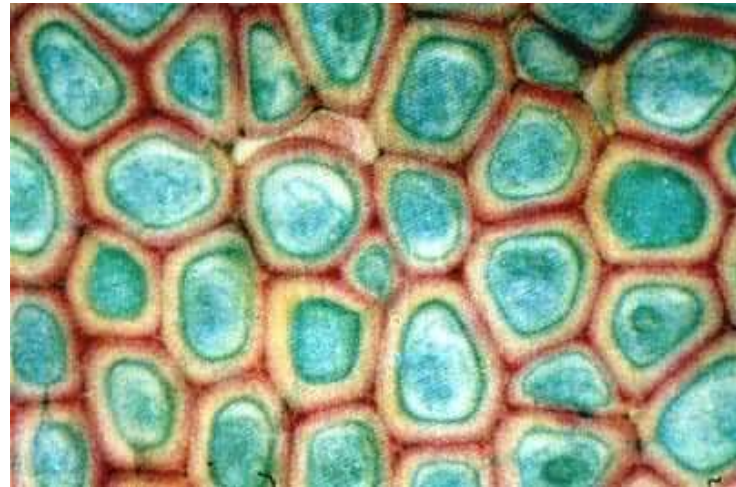 - books/dcpr/example/cleve/vshow.m

Natural Examples of Voronoi Diagrams (1/2)

Natural Examples of Voronoi Diagrams (2/2)

Characteristics of KNNC

- ## Strengths of KNNC   Quiz!
  - Intuitive
  - No computation for model construction
- ## Weakness of KNNC
  - Massive computation required when dataset is big
  - No straightforward way
    - To determine the value of K
    - To rescale the dataset along each dimension

Preprocessing of Feature Normalization

- **Z normalization** or **z score**
  - To have zero mean and unit variance along each feature
- **Range normalization**
  - To have a specific range, such as [0, 1], along each feature

Quiz!

Let $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ be the values of a specific feature of a dataset

Z normalization:
$$\widehat{x}_i = \frac{x_i - \mu}{\sigma}, \text{with } \mu \text{ and } \sigma^2 \text{ being the sample mean}$$
$$\text{and sample variance of } \mathbf{x} \text{ respectively}$$
Range normalization:
$$\hat{x}_i = \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \text{ to have a range of } [0, 1]$$
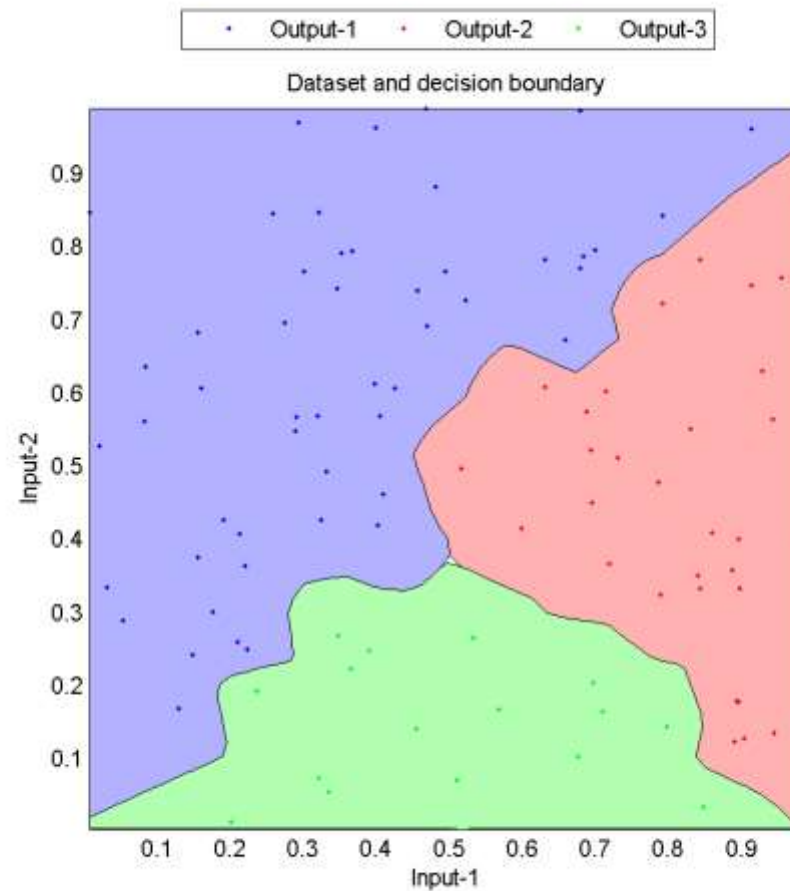
Variants for KNNC

○ Many variants of KNNC:

- Nearest prototype classification [Quiz!]
  - ○ Single prototype for each class ➜ Use "mean" or "average"
  - ○ Several prototypes for each class ➜ Use "k-means clustering"
- Distance-weighted votes
- Edited nearest neighbor classification
- k+k-nearest neighbor

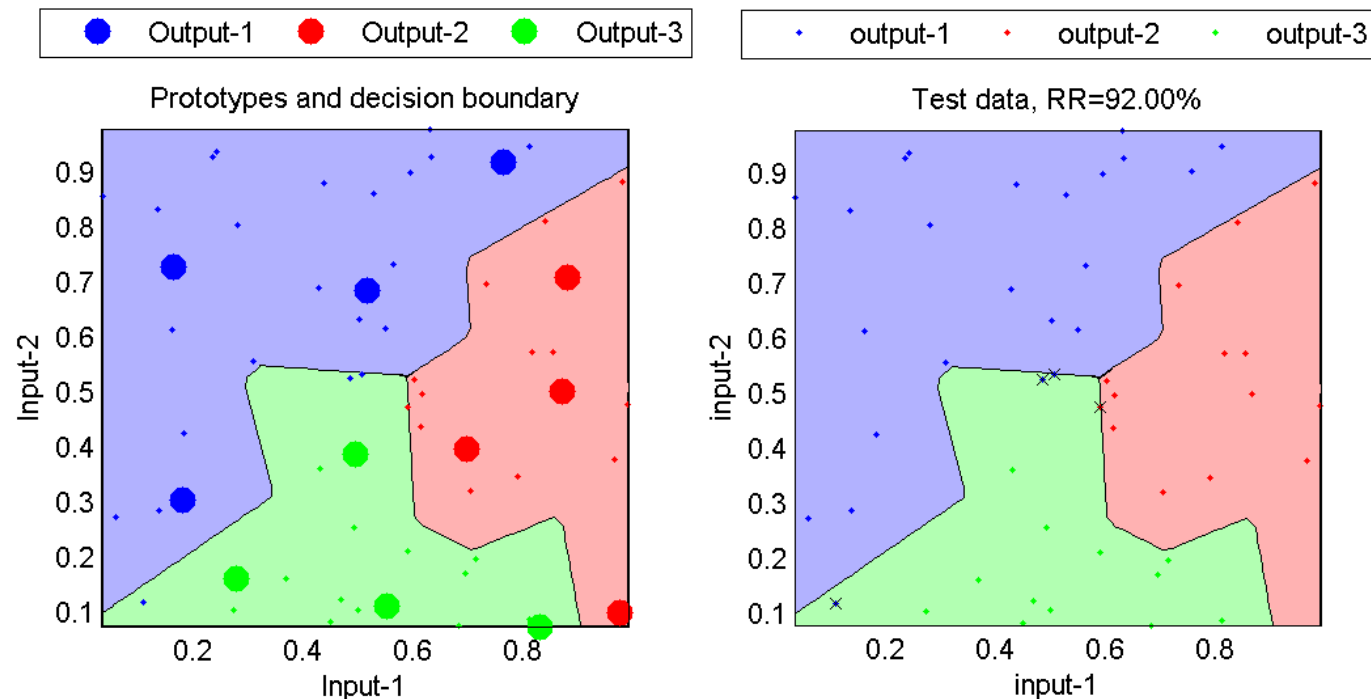# 1NNC Decision Boundaries

o 1NNC Decision boundaries

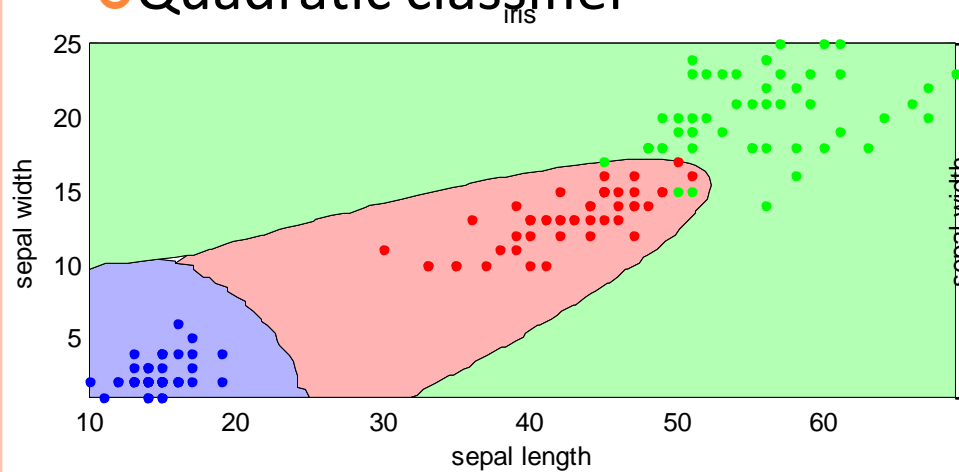# 1NNC Distance/Posterior as Surfaces and Contours
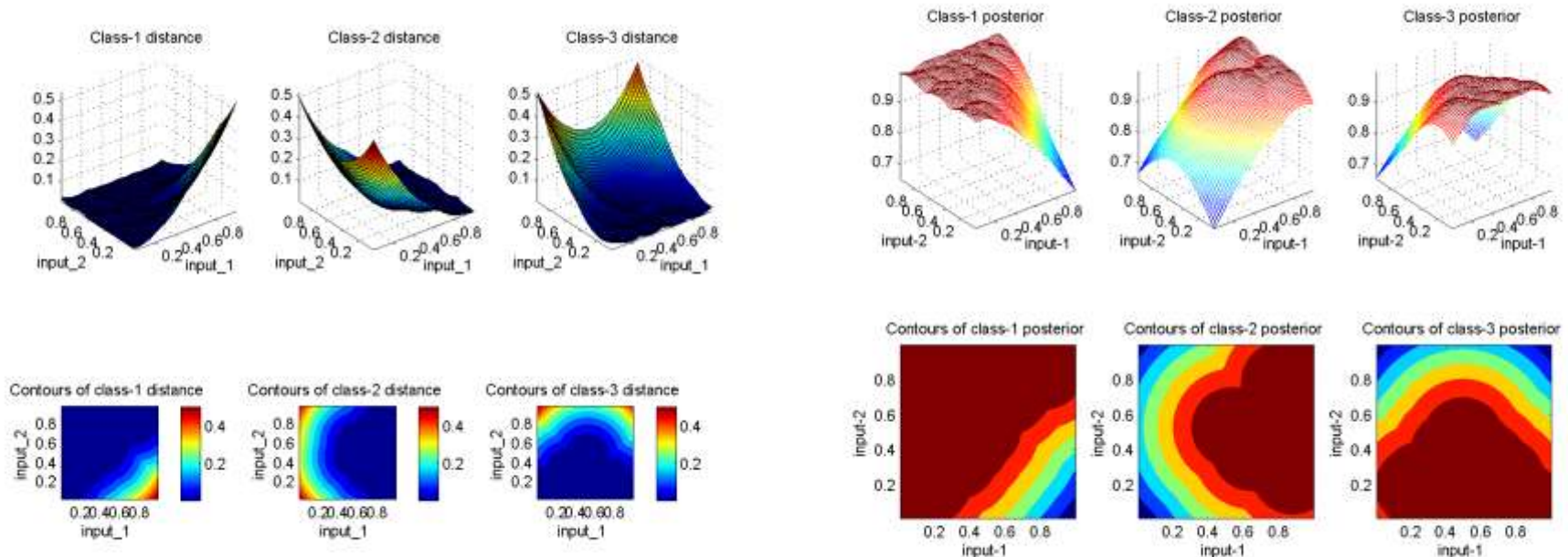
Using Prototypes in KNNC
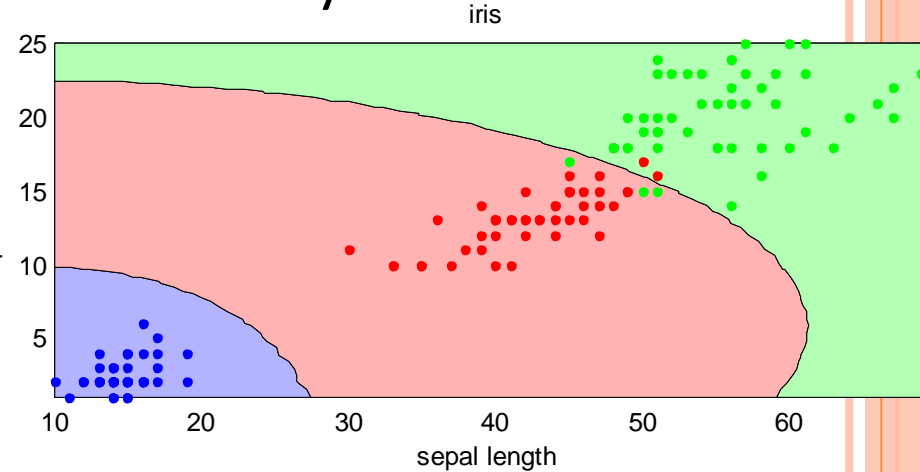
○ No. of prototypes for each class is 4.

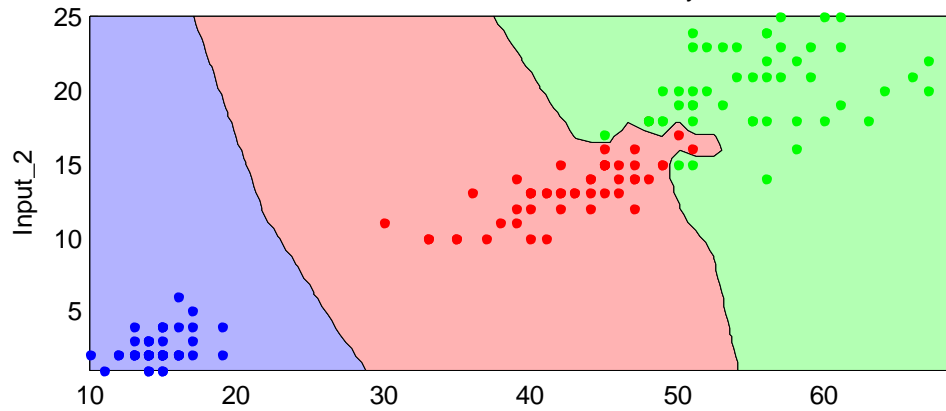# Decision Boundaries of Different Classifiers

○Quadratic classifier

○Naive Bayes classifier

○1NNC classifier

Exercise: KNNC Decision Boundary

○ Given 6 samples of two classes as shown below, plot the decision boundary based on KNNC with k=1.

Exercise: Nearest Prototype Classifier

- If we want to use the nearest prototype classifier instead of KNNC, what methods be used to find the prototype(s) for each class in the following two cases?

    - When the number of prototype is 1 for each class.

    - When the number of prototype is more than 1 for each class.