Pattern Classification using Rectified Nearest Feature Line Segment

Hao Du[1] and Yan Qiu Chen[1]

[1]Department of Computer Science and Engineering
Fudan University

ICNC'05 - FSKD'05, Aug.27th, 2005

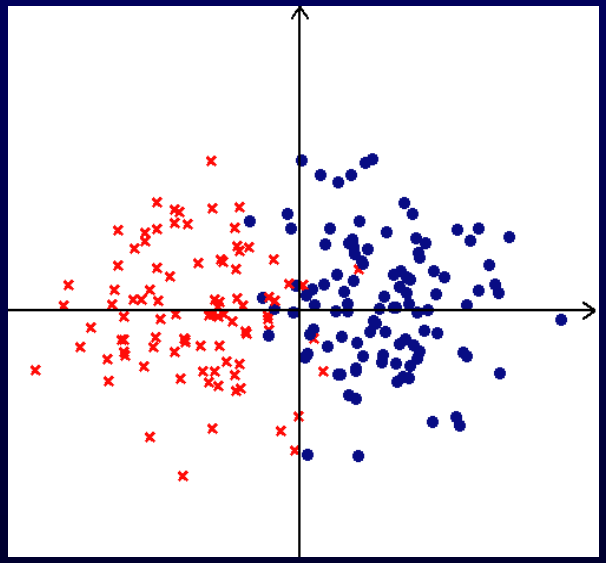
Pattern classifiers

- Nearest Neighbor (NN)
  - T.M. Cover and P.E. Hart, 1967
  - Simplest
  - Effective
  - Most popular

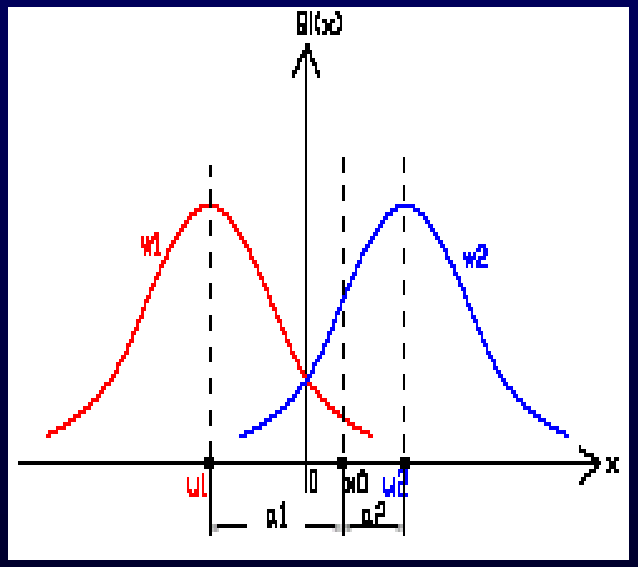
- The proposed method : RNFLS
  - A classifier to improve NN
  - by an average of 3% correct classification rate

Problem of Nearest Neighbor

- Risks occur in the intersection area



Sample Distribution

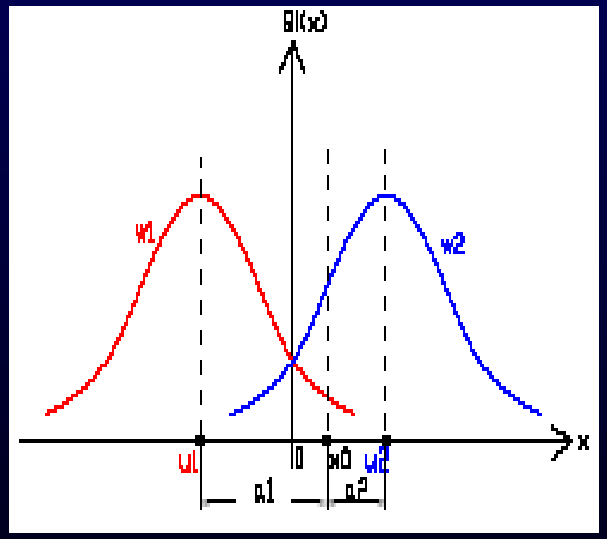


Density Function

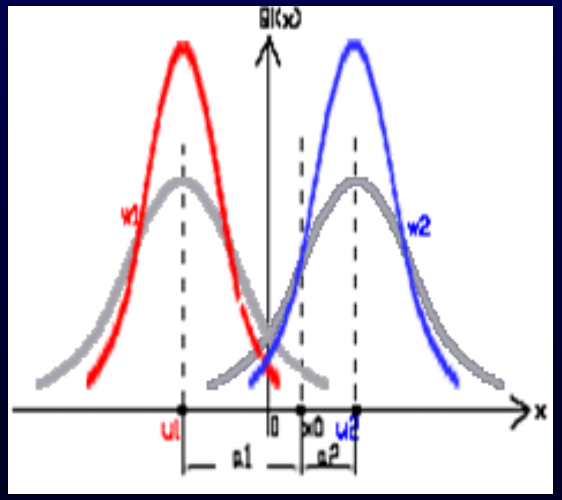
Motivation : Centralization

- How to improve NN in the intersection area?

> Centralizing the density function



Centralizing



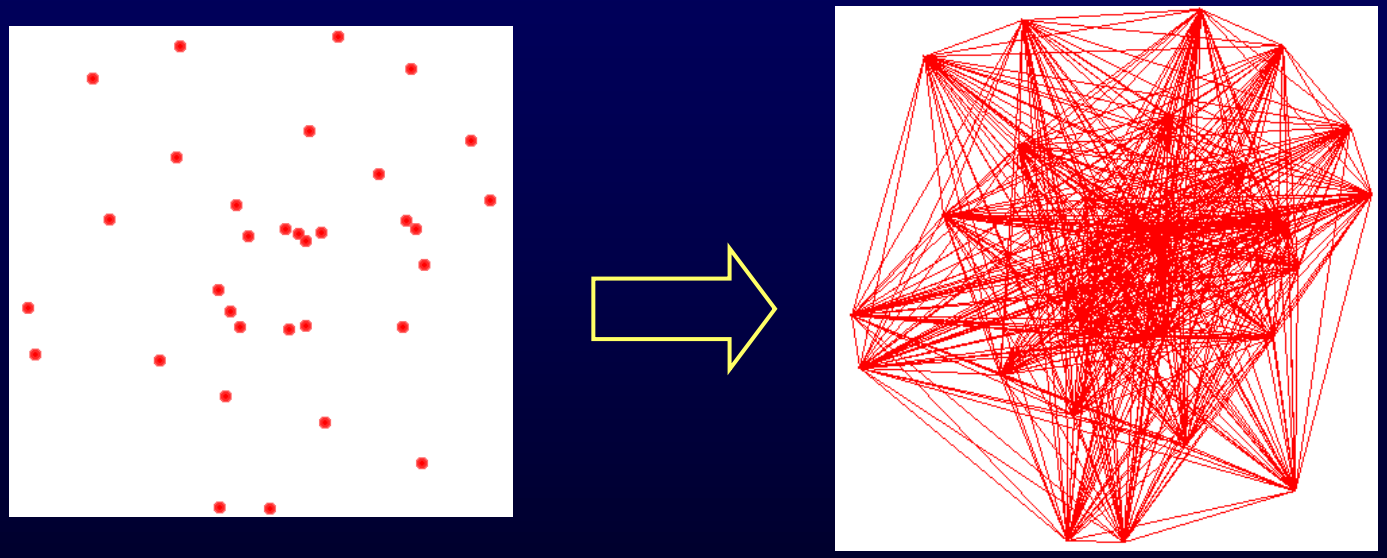
History: 'Feature Line'

- From NN -> NFL -> RNFLS
  - Nearest Neighbor (NN)
    - T.M. Cover and P.E. Hart,1967
  - Nearest Feature Line (NFL)
    - S.Z. Li and J. Lu, 1999
  - Rectified-NFL-Segment (RNFLS)
    - H. Du and Y.Q. Chen, this Paper

- NFL & RNFLS support the original sample points to improve NN classifier

The Method (1/2)

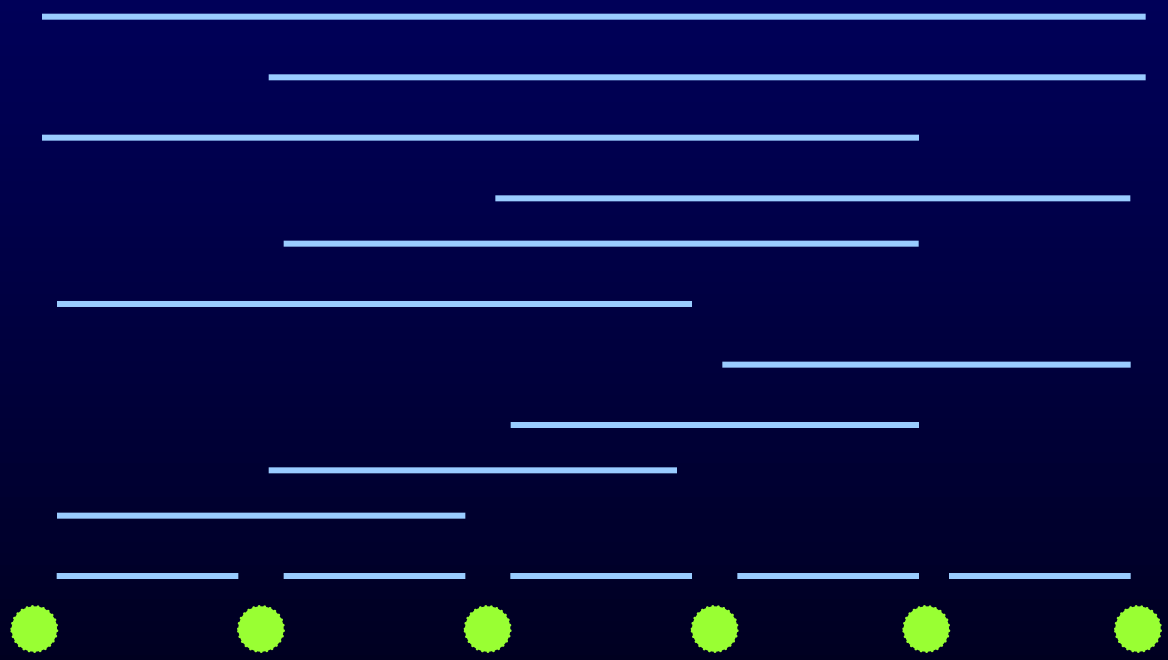
- Using line segments connect each pair of samples from one class



- Feature Line Segments support the original samples

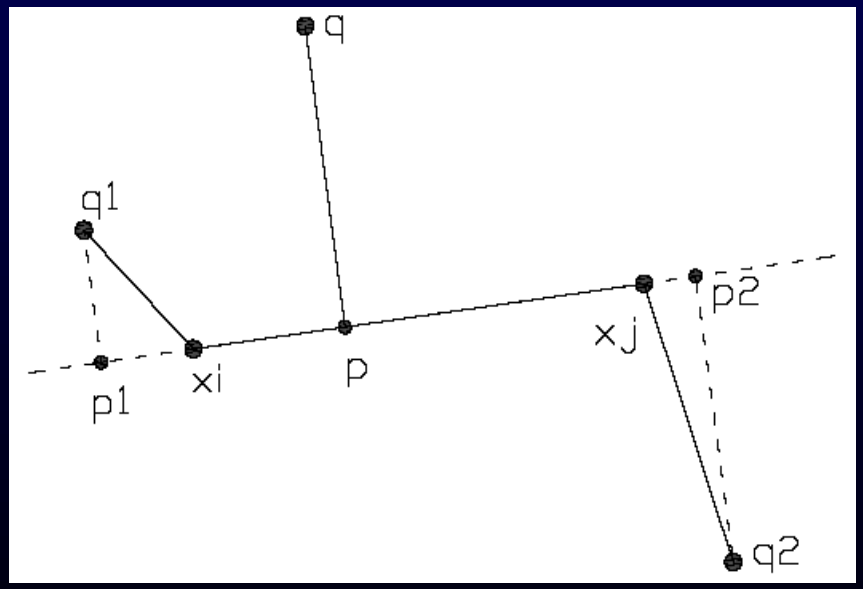
Why it works?

- 1D – example



The Method (2/2)

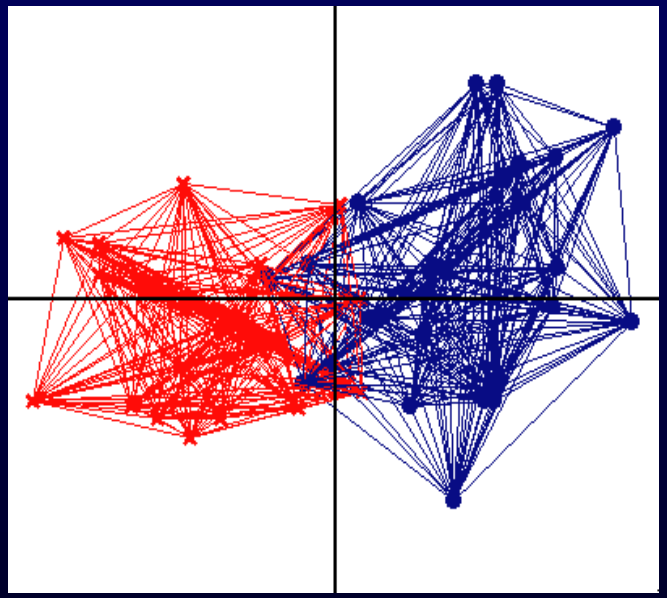
- Distance metric:
  - classifying by finding out the Nearest Feature Line Segment to the query point

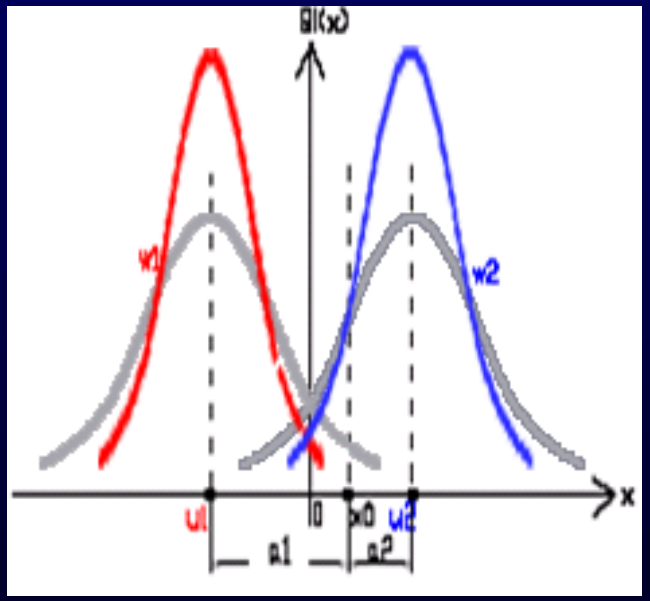


NFL-Segment

Two-Gaussian Example

- Sample distribution is centralized by NFL-Segment

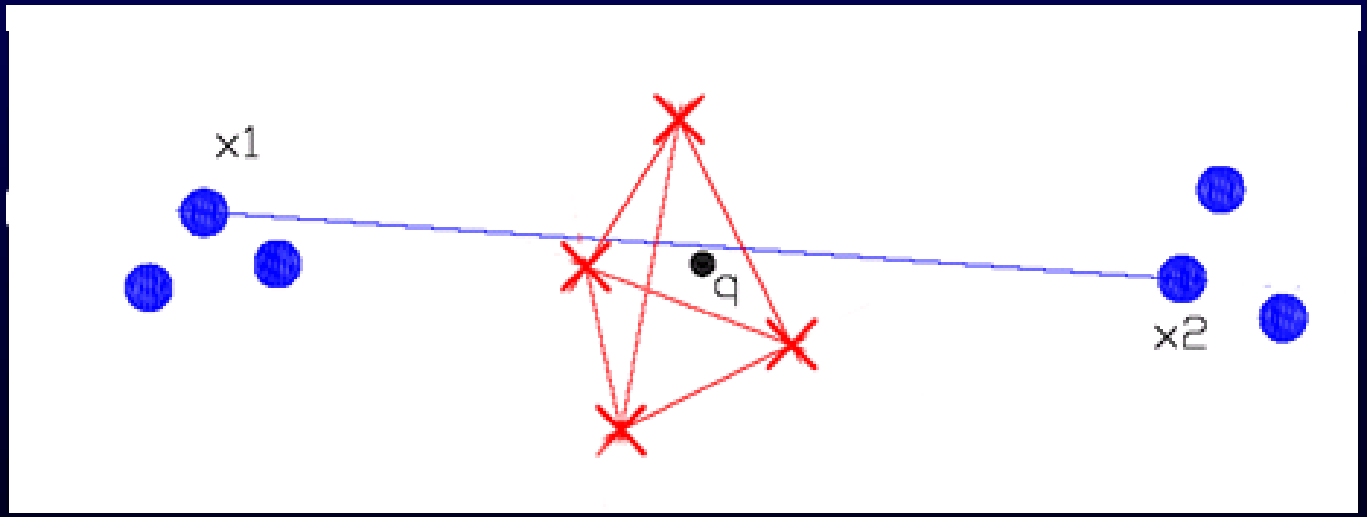




- Correct-classification-rate increases
  - 82% -> 86%

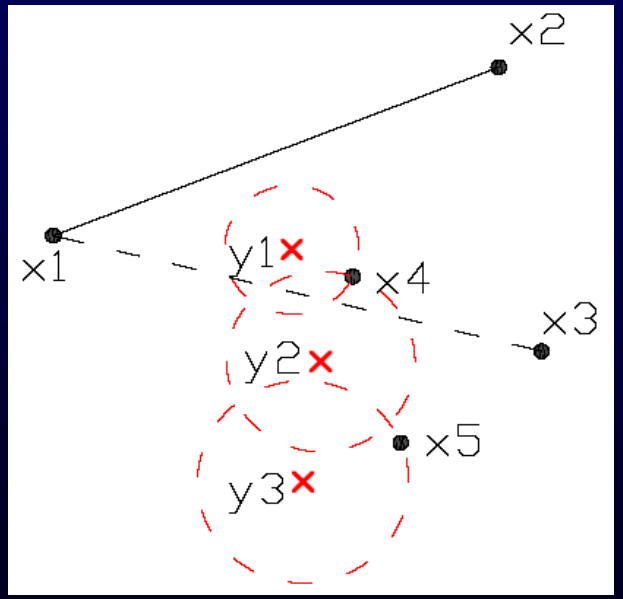
What if multi-center distribution?

- Problem:
  - Feature lines of one class may trespass other class area, causing decision errors



Rectified NFL-Segment

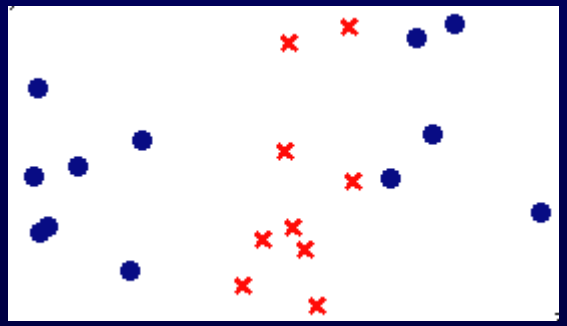
- Idea : remove all the feature lines trespassing the territory of other classes
- Territory
  - sample-territory
  - class-territory



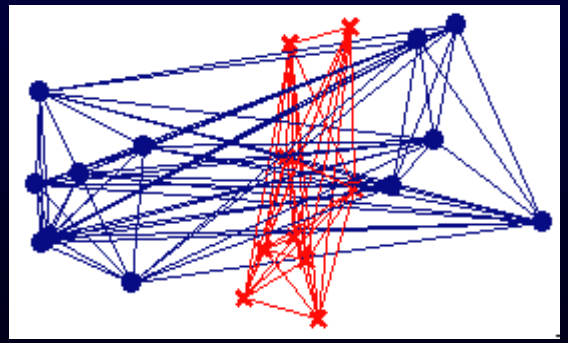
As shown in the picture, feature line segment X_1X_3 will be removed for it trespasses the territory of class “cross”

Review the whole process

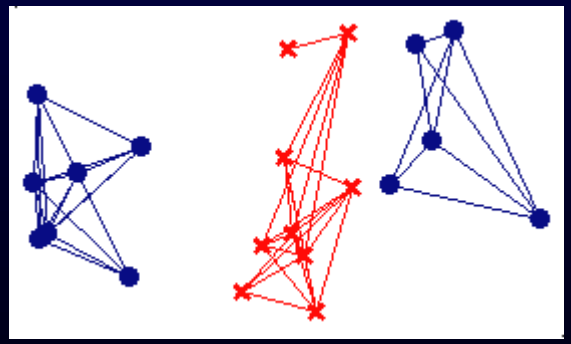
- Density Functions are centralized in different areas



(a) Nearest Neighbor



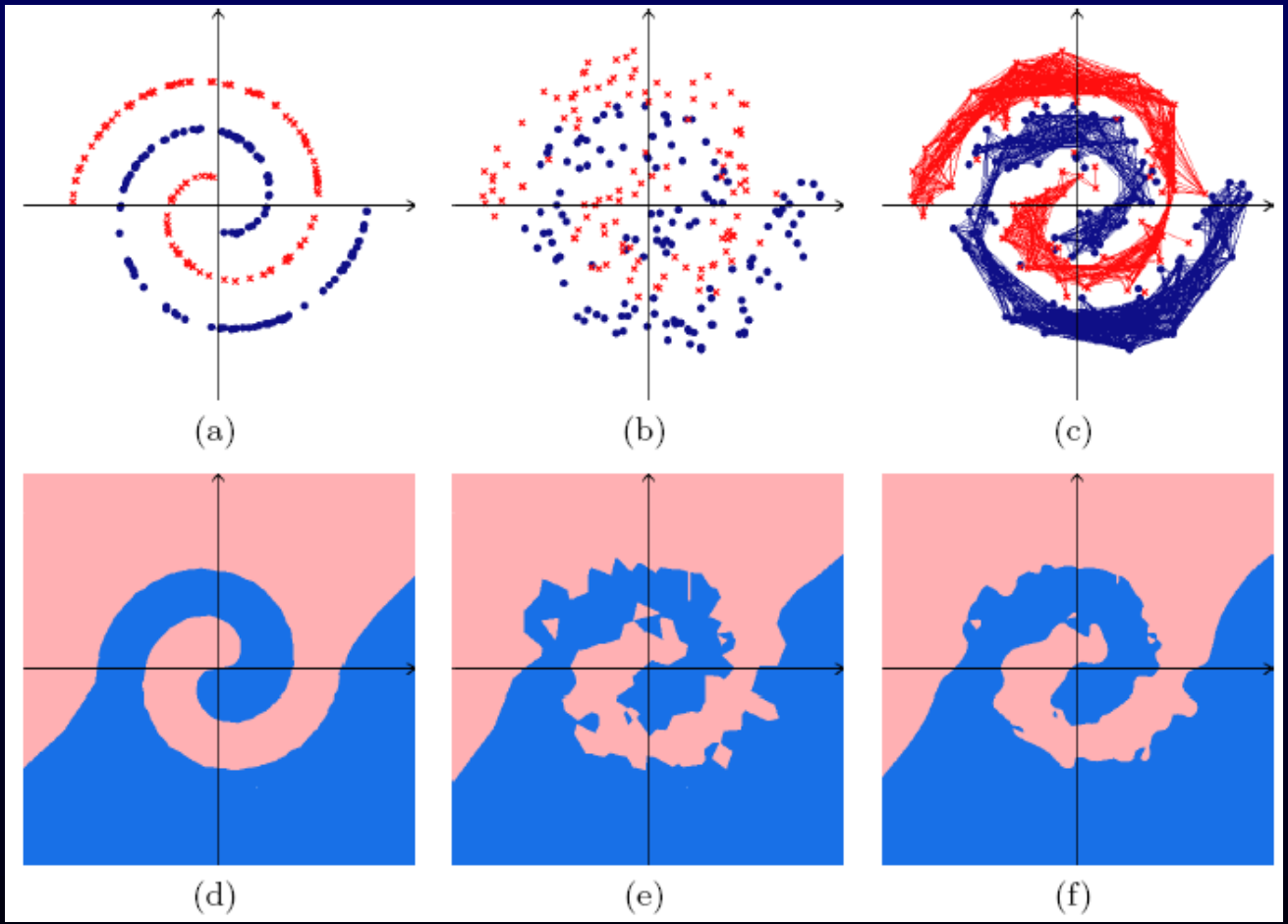
(b) NFL-Segment



(c) Rectified-NFL-Segment

Two-spiral Experiments

- Centralization property helps to optimize the decision region



Experiments: UCI-datasets

- Real-World Problems
  - UCI datasets
  - CCR(%) by leave-one-out procedure

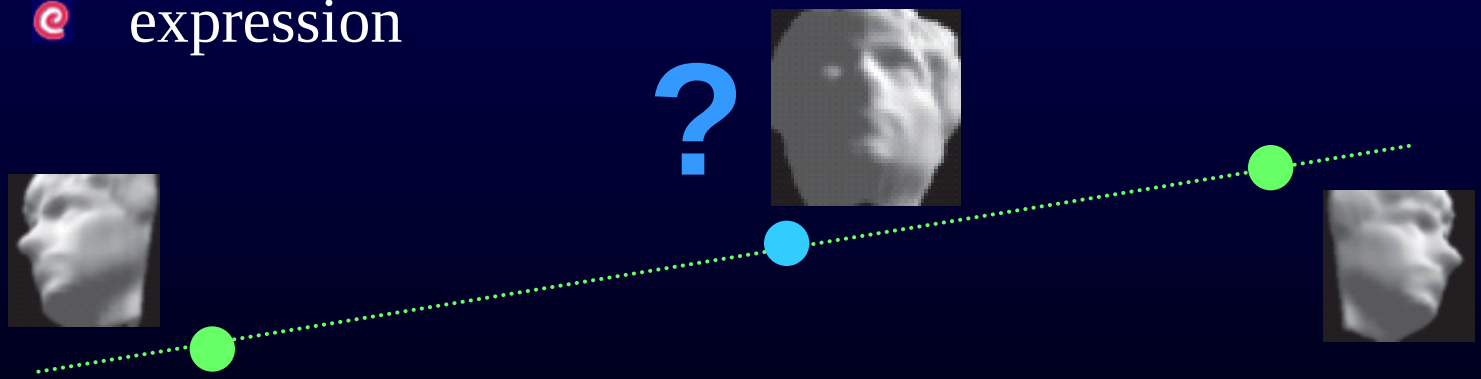
	Dataset	#Classes	#Instances	#Attributes	NN	3NN	NFL	NNL	RNFLS
1	hepatitis	2	80	19	**92.5**	91.3	91.3	76.3	91.3
2	iris	3	150	4	94.7	94.7	88.7	94.7	**95.3**
3	housing	6	506	13	70.8	73.0	71.1	67.6	**73.5**
4	pima	2	768	8	70.6	**73.6**	67.1	62.8	73.0
5	wine	3	178	13	95.5	95.5	92.7	78.7	**97.2**
6	bupa	2	345	6	63.2	65.2	63.5	57.4	**66.4**
7	ionosphere	2	351	34	86.3	84.6	85.2	87.2	**94.3**
8	wpbc	2	194	32	72.7	68.6	72.7	54.1	**75.8**
9	wdbc	2	569	30	95.1	96.5	95.3	64.0	**97.2**
10	glass	6	214	9	70.1	72.0	66.8	65.4	**72.0**

Interesting property of Feature Line(1/3)

- S.Z. Li (1999), Nearest Feature Line:

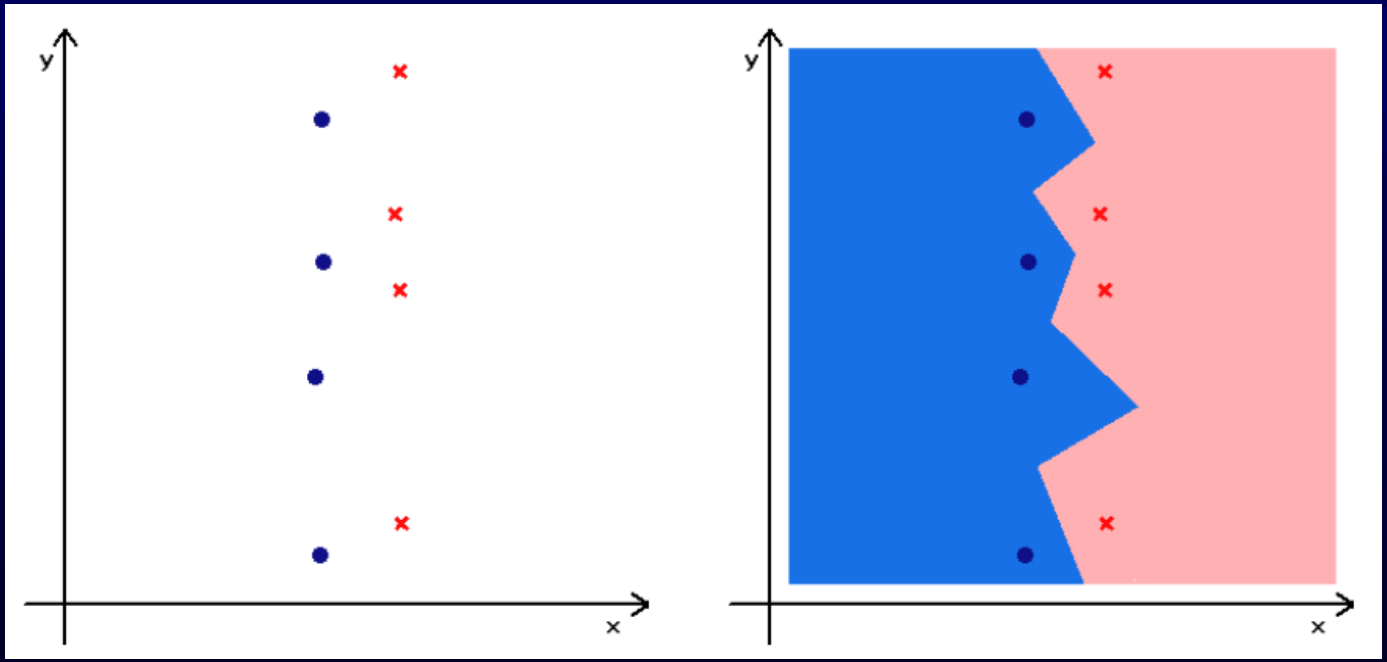
In Face-Recognition problem, Feature Line linearly expand two prototypes, approximating variants of two prototypes in

- pose
- illumination
- expression



Interesting property of Feature Line(2/3)

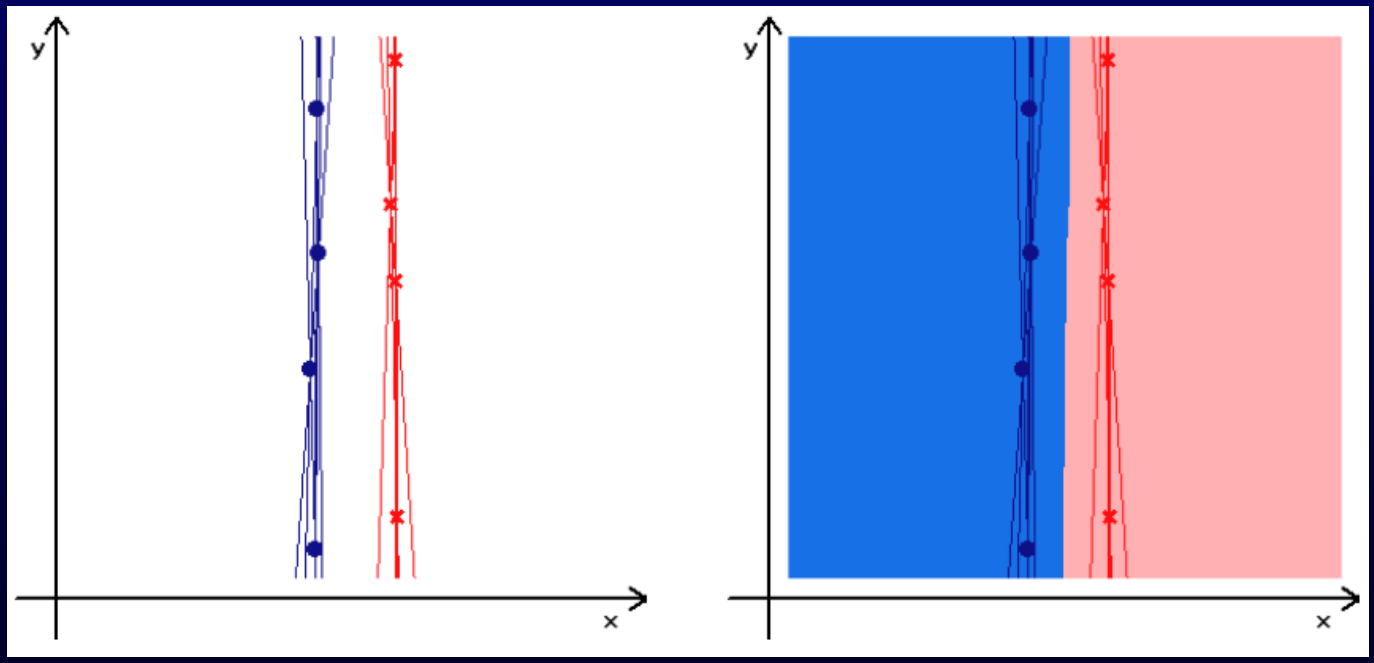
- Another example:



(a) Original Samples (b) Decision Region by NN

Interesting property of Feature Line(3/3)

- Another example:



(c) Feature Lines (d) Decision Region by NFL

Q & A

Thank you!