CIS501 – Lecture 4

Woon Wei Lee
Fall 2013, 10:00am-11:15am,
Sundays and Wednesday

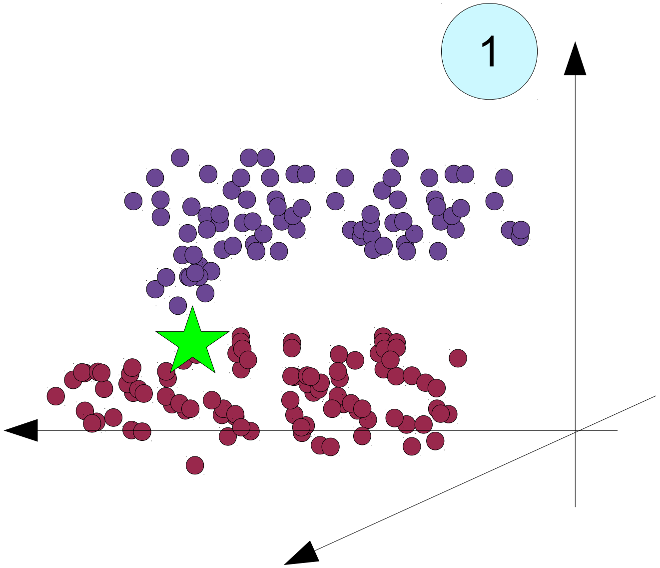
Masdar INSTITUTE

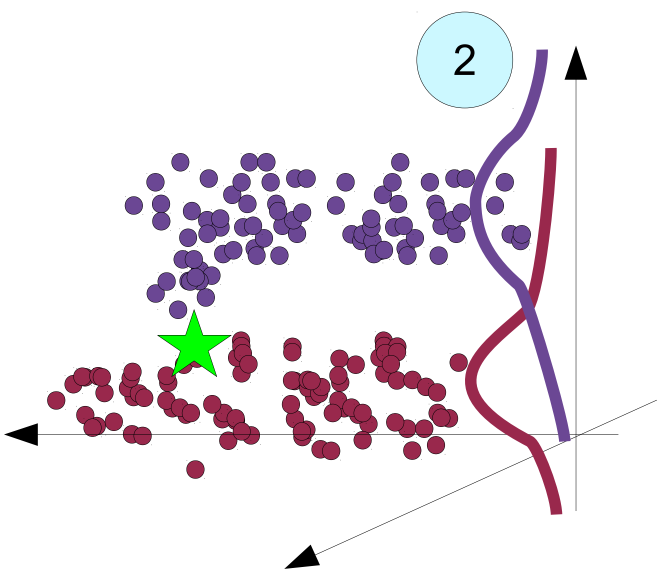
For today:

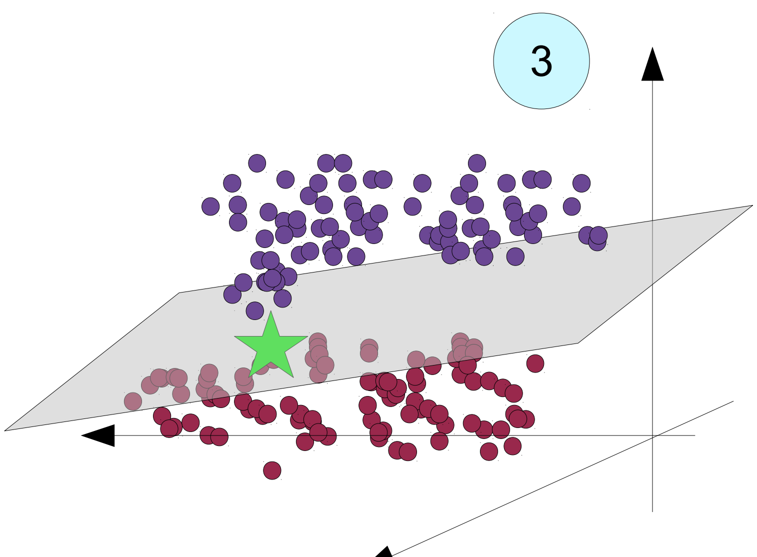
- Administrative stuff and discussions
  - Presentations!
- Classification intro:
  - Discriminative vs Generative classification
    - Density Estimation
    - Bayes decision rule

Classification: Discriminative vs Generative

- **The challenge**
  - Distinguish between data from two “classes”
- Geometrical perspective shown on right (fig.1).
  - There is training data (small coloured points) and an “unseen” instance (star)
- **Two approaches:**
  1. Determine the class conditional distributions → assign point to most likely class (fig. 2)
  - Identify discriminative direction or features (fig. 3)

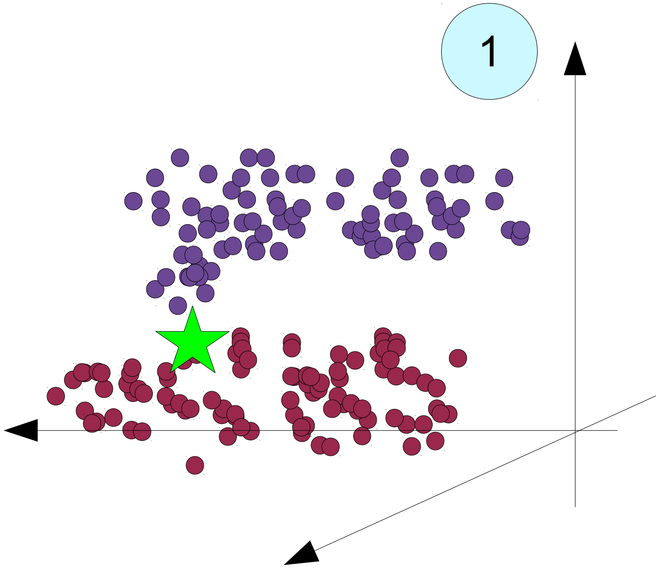


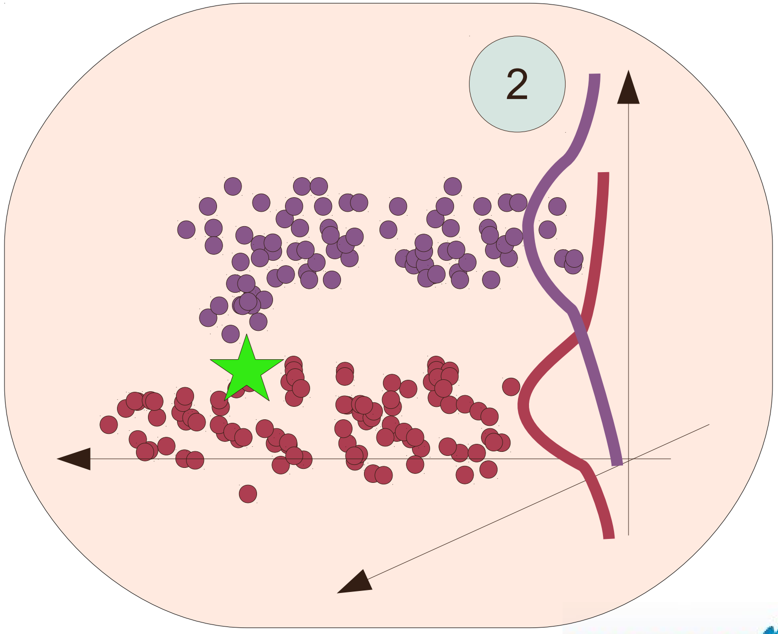


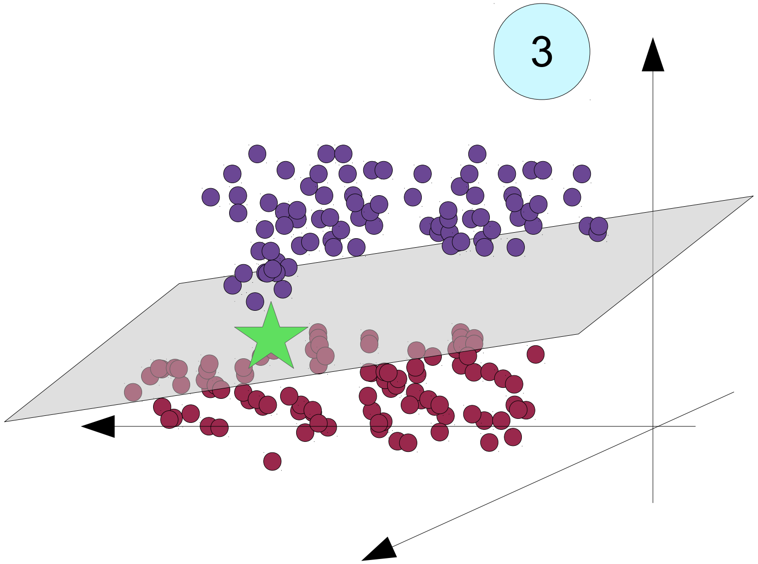


Classification: Discriminative vs Generative

- **The challenge**
  - Distinguish between data from two “classes”
- Geometrical perspective shown on right (fig.1).
  - There is training data (small coloured points) and an “unseen” instance (star)
- **Two approaches:**
  1. Determine the class conditional distributions → assign point to most likely class (fig. 2)
  - Identify discriminative direction or features (fig. 3)





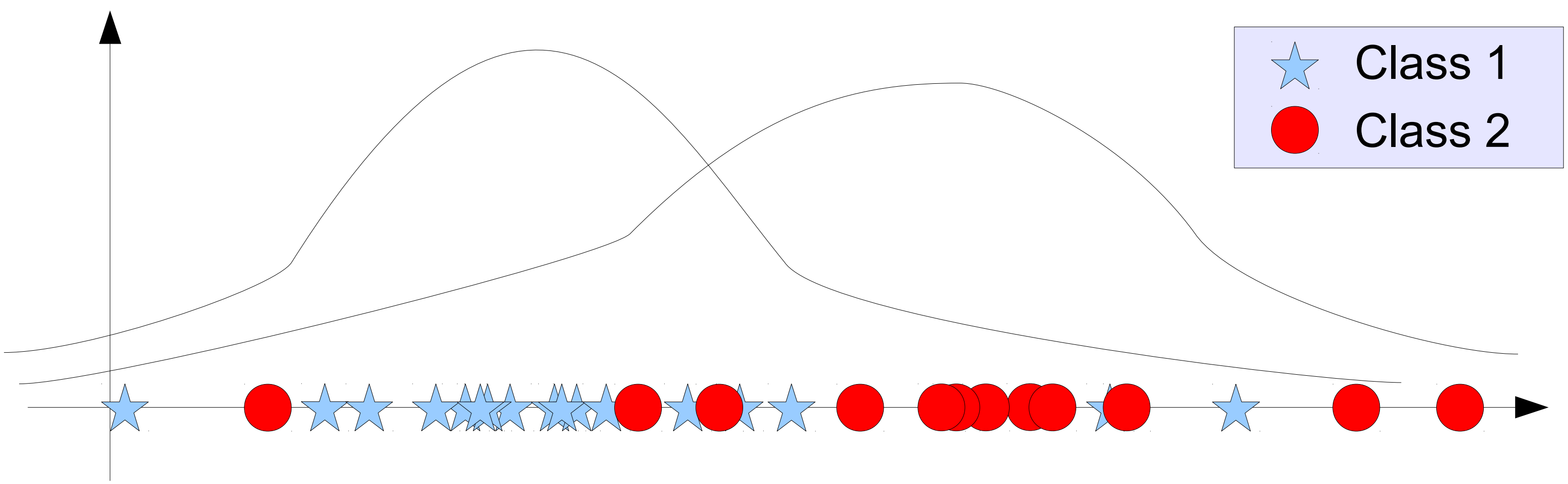


The probabilistic perspective:
Generative classifiers

- **Underlying concepts:**
  - Notion that the data is a manifestation of an underlying *generator*
  - Characterized by its *distribution* **p*(x)***
  - Classification of new points based on goodness-of-fit to this assumed generator
- **Steps in modeling:**
  1. Model selection
  2. Density estimation of the data
  3. Classification of new data
- **Models covered:**
  - *k*NN
  - Naive Bayes Classifier

Density estimation

- How can we estimate the densities p(x)?
- Start with data that is pre-“labelled” as belonging to either Class 1 (c_1) or 2 (c_2) $\rightarrow$ “**Training data**”
- The instances corresponding to the individual classes are used in isolation to fix p($x|c_i$)



(simple) parametric example: **Gaussian distribution**

- Two-class training data:

c_1	**0**	**1.5**	**2**	**2.5**	**4**
c_2	**3.5**	**5**	**7**	**8**	**10**

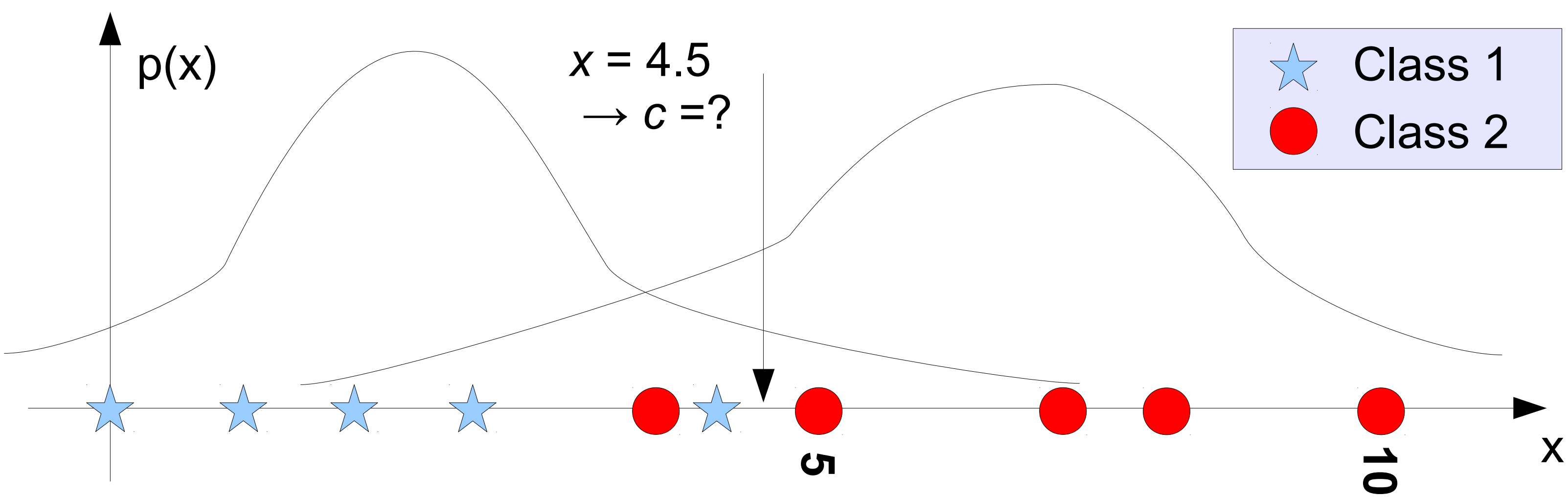
- Model each class as a gaussian..

$$p(x|c_1)=N(1.8,2.575)\ ;\ \ p(x|c_2)=N(6.7,6.45)$$

- x=4.5 (identify class)

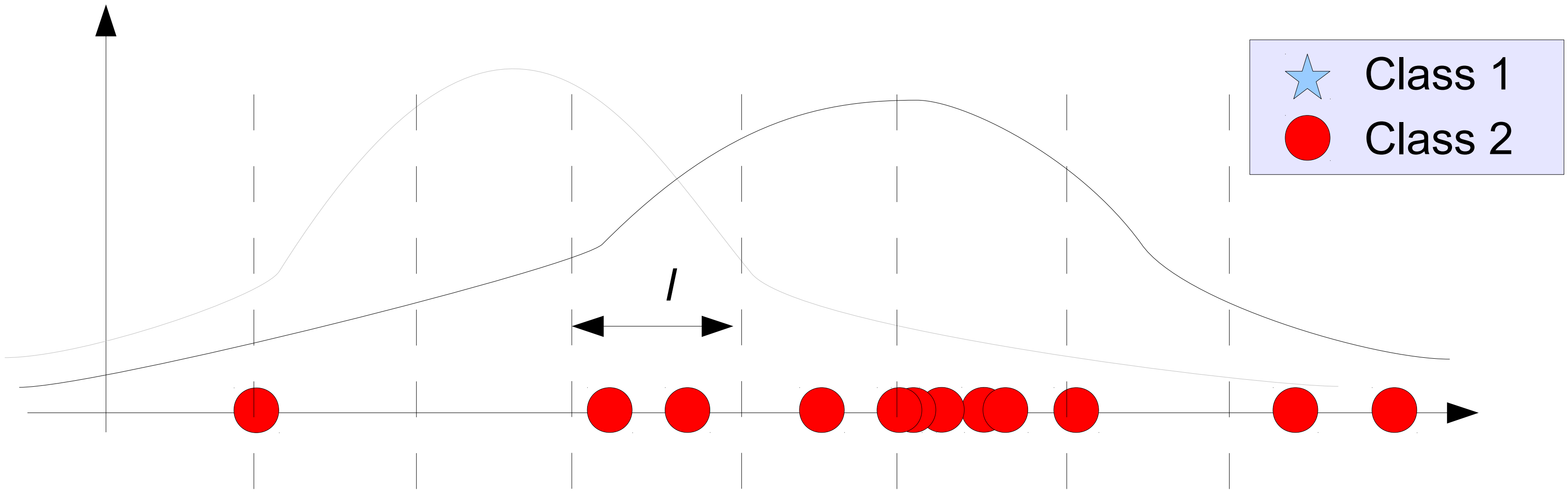
$$p(x|c_1)=\frac{1}{\sqrt{2\pi 2.575}}e^{-\left[\frac{(4.5-1.8)^2}{2\times 2.575}\right]}=0.060362\ \ ,\ \ p(x|c_2)=0.10794$$

$\Rightarrow$ c $\rightarrow$ c_2 (Congrats, your first classifier!)



Masdar INSTITUTE

Non-parameteric estimation

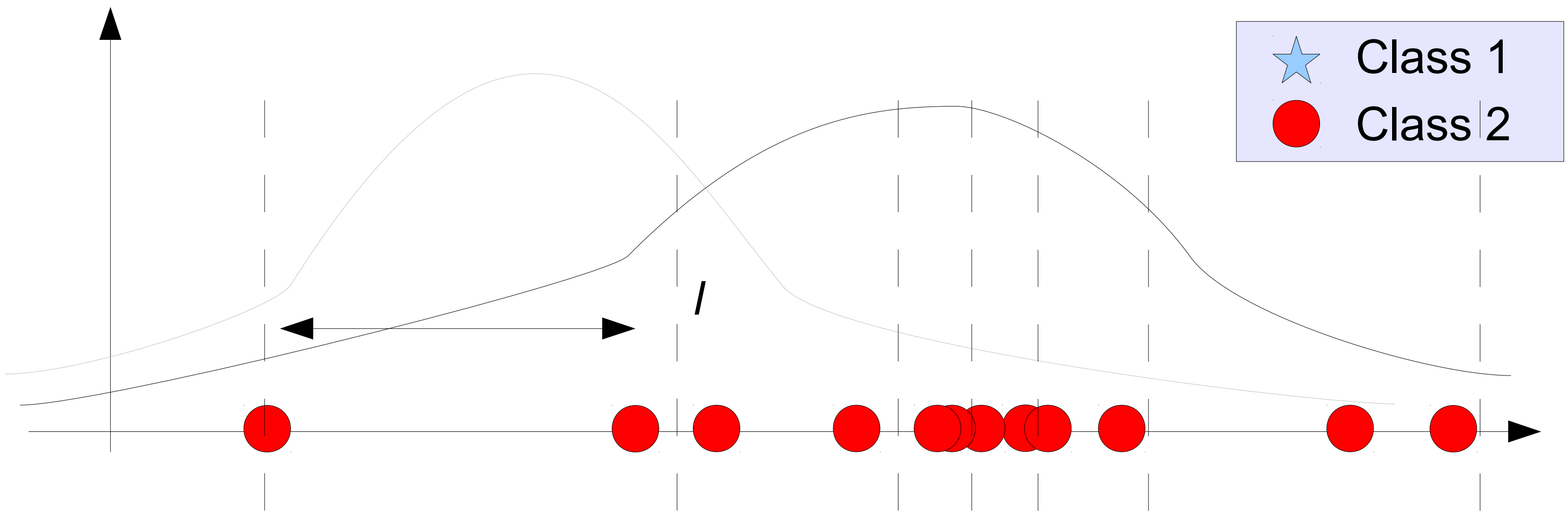


- Crudest method: histogram method (illustrated above)
- Divide axis into uniformly sized bins, calculate:

$$p\left(x|c_2\right)=\frac{n}{Nl}\propto\frac{1}{6}$$

- Notes:
  - Conceptually simple/easy to implement
  - **Q: What are the disadvantages?**

k-NN approach



- Still based on:

$$p(x|c_2) = \frac{k}{Nl} = \frac{1}{6l}$$

- i.e. only *l* changes
- Allows the size of the sampling area to change w.r.t. the data distribution
- Sparse areas of the axes can still yield non-zero probabilities
- **Question: What does changing the value of *k* do?**

Masdar INSTITUTE

(Refresher) **Bayes Theorem**

- Bayes theorem is given by:

$$p(c|x) = \frac{p(c, x)}{p(x)}$$
$$= \frac{p(x|c)\,p(c)}{p(x)}$$

- Specialized terms in Bayesian Analysis:
  - *c* - The model or property to be inferred
  - *x* -The “observations”
  - *p(x|c)* – The “Likelihood”
  - *p(c)* – The “prior”
  - *p(c|x)* – The “posterior”
  - *p(x)* – The “evidence”

Thomas Bayes
1702-1761

Bayes decision rule

- In the case of classification, "*c*" denotes the category or class from which the data was sampled
- In general, the classification problem is as follows:
  - Given a particular observation, *x* and *n* potential classes, determine the class which satisfies:

$$c = \underset{\forall i \in \{1,2,\ldots,n\}}{argmax_i}\; p(c_i|x)$$

- *p(x)* is independent of the class, and..
- .. *p(c)* is frequently assumed to be the same for all classes.
- In which case, the likelihood term is interchangeable with the posterior term above.