

Only covered supervised learning $X \rightarrow \mathbb{R}$ regression
 $X \rightarrow \{0, 1, \dots, k\}$ classification

Training data included labels

Clustering

K-means

Machine Learning – CSE546
 Carlos Guestrin
 University of Washington

November 4, 2014
 ©Carlos Guestrin 2005-2014

Clustering images

Set of Images

given no labels

organize data into themes

beaches

flowers

C₁

C₂

C₃

C₄

C₅

©Carlos Guestrin 2005-2014 [Goldberger et al.] 2

Clustering web search results

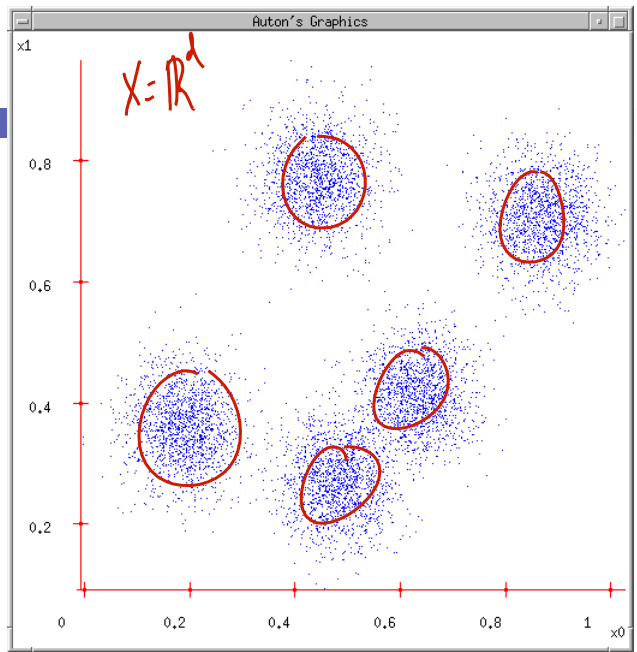
The screenshot shows the Clusty search interface for the query 'race'. On the left, a sidebar lists various clusters such as 'Car (2)', 'Race cars (7)', 'Photos, Races Scheduled (5)', 'Game (4)', 'Track (3)', 'Nascar (3)', 'Equipment And Safety (2)', 'Other Topics (7)', 'Photos (22)', 'Game (14)', 'Selfies (13)', 'Team (18)', 'Human (8)', 'Classification Of Human (2)', 'Statement, Evolved (2)', 'Other Topics (4)', 'Weekend (8)', 'Ethnicity And Race (7)', 'Race for the Cure (8)', and 'Race Information (8)'. The main content area displays a list of search results, each with a title and a brief description. A red circle highlights the 'Human' cluster in the sidebar and the first search result, 'Race (classification of human beings) - Wikipedia, the free encyclopedia'.

Cluster Human contains 8 documents.

- Race (classification of human beings) - Wikipedia, the free encyclopedia**
The term race or racial group usually refers to the concept of dividing humans into populations or groups on the basis of various sets of characteristics. The most widely used human racial categories are based on visible traits (especially skin color, cranial or facial features and hair texture), and self-identification. Conceptions of race, as well as specific ways of grouping races, vary by culture and over time, and are often controversial for scientific as well as social and political reasons. History - Modern debates - Political and ...
- Race - Wikipedia, the free encyclopedia**
General: Racing competitions The **Race** (yachting race), or La course du millénaire, a no-rules round-the-world sailing event; **Race** (biology), classification of flora and fauna; **Race** (classification of human beings) **Race** and ethnicity in the United States Census, official definitions of "race" used by the US Census Bureau; **Race** and genetics, notion of racial classifications based on genetics. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** in molecular biology "Rapid ... General - Surnames - Television - Music - Literature - Video games
- Publications | Human Rights Watch**
The use of torture, unlawful rendition, secret prisons, unfair trials, ... Risks to Migrants, Refugees, and Asylum Seekers in Egypt and Israel ... In the run-up to the Beijing Olympics in August 2008, ...
- Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich, Frank Miele: Books ...**
Amazon.com: **Race: The Reality Of Human Differences: Vincent Sarich, Frank Miele:** Books ... From Publishers Weekly Sarich, a Berkeley emeritus anthropologist, and Miele, an editor ...
- AAPA Statement on Biological Aspects of Race**
AAPA Statement on Biological Aspects of Race ... Published in the American Journal of Physical Anthropology, vol. 101, pp 569-570, 1996 ... PREAMBLE As scientists who study human evolution and variation, ...
- race, Definition from Answers.com**
race n. A local geographic or global human population distinguished as a more or less distinct group by genetically transmitted physical ...
- Dopefish.com**
Site for newbies as well as experienced Dopefish followers, chronicing the birth of the Dopefish, its numerous appearances in several computer games, and its eventual take-over of the human race. Maintained by Mr. Dopefish himself, Joe Siegler of Apogee Software.

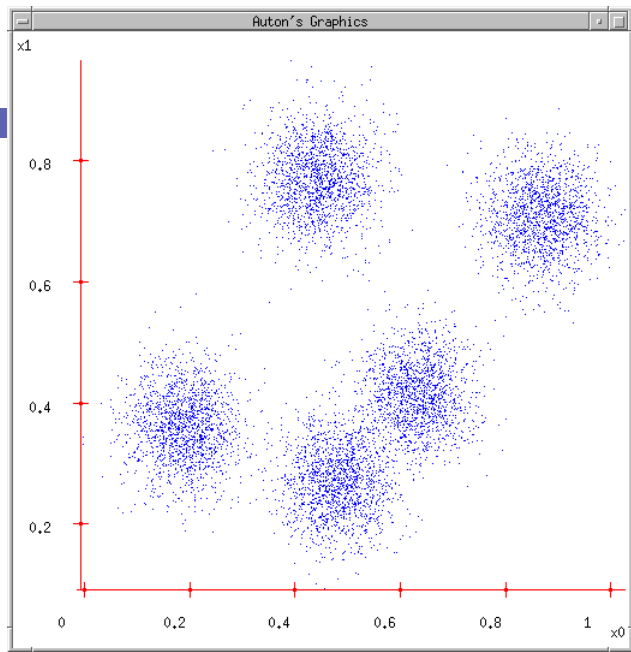
©Carlos Guestrin 2005-2014 3

Some Data



K-means

1. Ask user how many clusters they'd like.
(e.g. $k=5$)

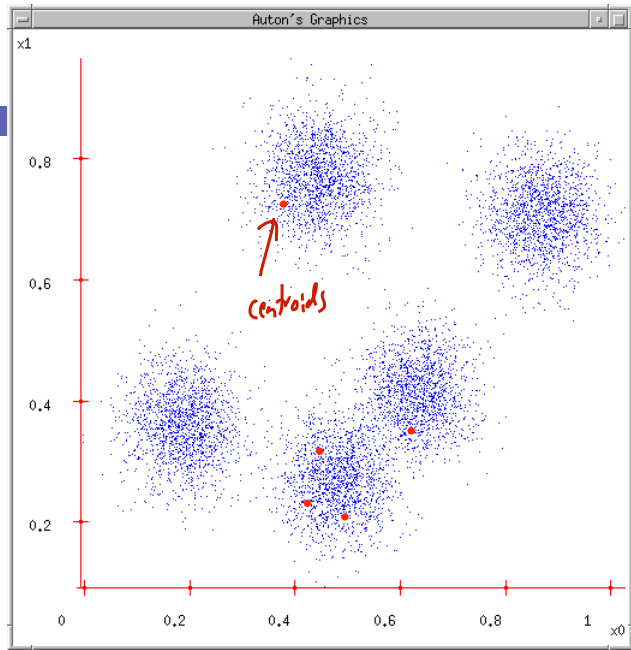


©Carlos Guestrin 2005-2014

5

K-means

1. Ask user how many clusters they'd like.
(e.g. $k=5$)
2. Randomly guess k cluster Center locations

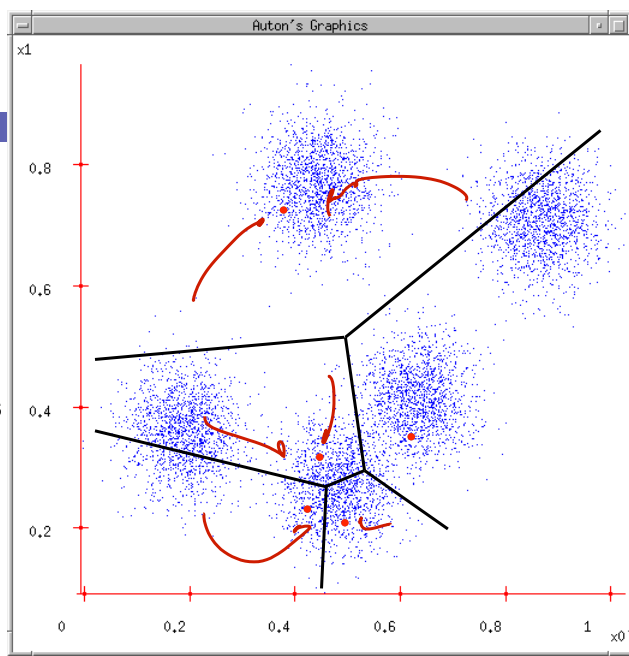


©Carlos Guestrin 2005-2014

6

K-means

1. Ask user how many clusters they'd like. (e.g. $k=5$)
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to. (Thus each Center "owns" a set of datapoints)

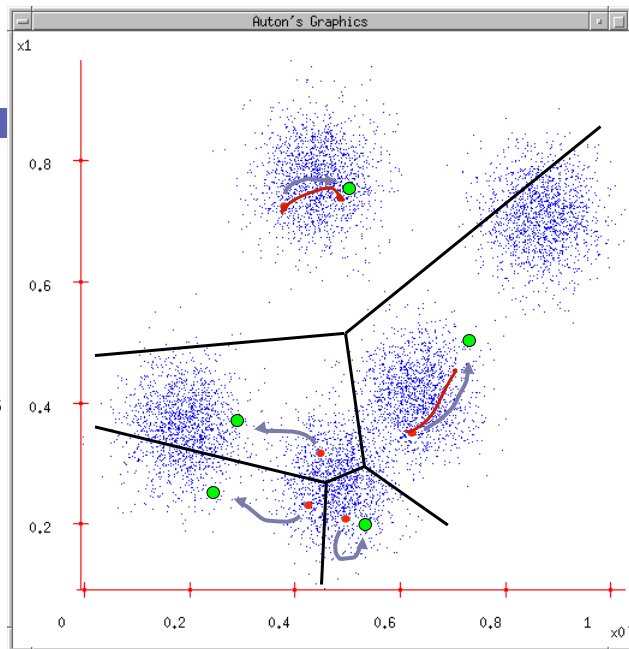


©Carlos Guestrin 2005-2014

7

K-means

1. Ask user how many clusters they'd like. (e.g. $k=5$)
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns

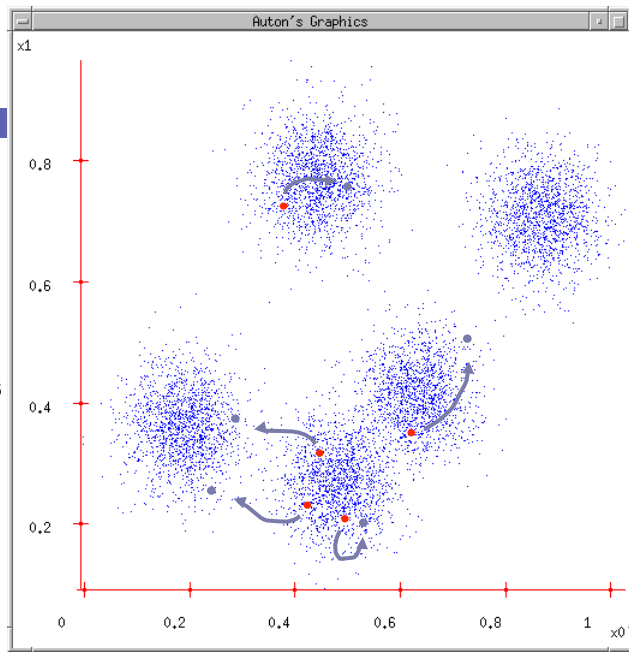


©Carlos Guestrin 2005-2014

8

K-means

1. Ask user how many clusters they'd like. (e.g. $k=5$)
2. Randomly guess k cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns...
5. ...and jumps there
6. ...Repeat until terminated!



©Carlos Guestrin 2005-2014

9

K-means

- Randomly initialize k centers

$\mu^{(0)} = \mu_1^{(0)}, \dots, \mu_k^{(0)}$ *intention*

d-dim vectors

or "smartly"

Repeat until convergence: no point changes) cluster membership

- **Classify:** Assign each point $j \in \{1, \dots, N\}$ to nearest center:

$C^{(t)}(j) \leftarrow \arg \min_i \|\mu_i^{(t)} - x_j\|^2$

fix μ , opt C

- **Recenter:** $\mu_i^{(t+1)}$ becomes centroid of its point:

$\mu_i^{(t+1)} \leftarrow \arg \min_{\mu} \sum_{j: C^{(t)}(j)=i} \|\mu - x_j\|^2$

sum of points in cluster i

$\mu_i^{(t+1)} = \frac{\sum_{j: C^{(t)}(j)=i} x_j}{|\{j: C^{(t)}(j)=i\}|}$

- Equivalent to $\mu_i \leftarrow$ average of its points!

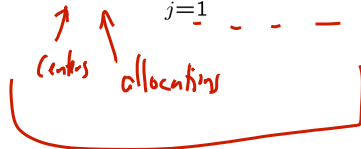
©Carlos Guestrin 2005-2014

10

What is K-means optimizing?

- Potential function $F(\mu, C)$ of centers μ and point allocations C :

$$\square F(\mu, C) = \sum_{j=1}^N \|\mu_{C(j)} - x_j\|^2$$



(coordinate descent!!)

- Optimal K-means:

$$\square \min_{\mu} \min_C F(\mu, C)$$

← unfortunately, simultaneous optimization is very hard

Does K-means converge??? Part 1

- Optimize potential function:

$$\min_{\mu} \min_C F(\mu, C) = \min_{\mu} \min_C \sum_{i=1}^k \sum_{j: C(j)=i} \|\mu_i - x_j\|^2$$

- Fix $\mu^{(t)}$, optimize C

← Sum over points in each cluster

$$\min_{C: C(1), \dots, C(j), \dots, C(N)} \sum_{j=1}^N \|\mu_{C(j)}^{(t)} - x_j\|^2 = \min_{C(1) \dots C(j)} \min_{C(N)} \sum_{i=1}^N \|\mu_{C(i)}^{(t)} - x_j\|^2$$

$$= \sum_{j=1}^N \min_{C(j)} \|\mu_{C(j)}^{(t)} - x_j\|^2$$

indep. OPT problems

find closest center to x_j
 \Rightarrow "classification step"

Does K-means converge??? Part 2

- Optimize potential function:

$$\min_{\mu} \min_C F(\mu, C) = \min_{\mu} \min_C \sum_{i=1}^k \sum_{j: C(j)=i} \| \mu_i - x_j \|^2$$

- Fix $C^{(t)}$, optimize μ

$$\min_{\mu_1} \dots \min_{\mu_k} \sum_{i=1}^k \sum_{j: C(j)=i} \| \mu_i - x_j \|^2$$

$$= \sum_{i=1}^k \min_{\mu_i} \sum_{j: C(j)=i} \| \mu_i - x_j \|^2$$

recenter step

©Carlos Guestrin 2005-2014

13

Coordinate descent algorithms

$$\min_{\mu} \min_C F(\mu, C) = \min_{\mu} \min_C \sum_{i=1}^k \sum_{j: C(j)=i} \| \mu_i - x_j \|^2$$

- Want: $\min_a \min_b F(a, b)$
 - Coordinate descent:
 - fix a, minimize b
 - fix b, minimize a
 - repeat
 - Converges!!!
 - if F is bounded
 - to a (often good) local optimum
 - as we saw in applet (play with it!)
 - (For LASSO it converged to the global optimum, because of convexity)
- random restarts help

$$\begin{matrix} a, b & F(a, b) & & \\ \downarrow & \geq & & \\ a', b' & F(a', b') & & \\ \downarrow & \geq & & \\ a'', b'' & F(a'', b'') & & \end{matrix}$$

||||| F bounded
F ≥ 0

picking k:
- $F(\mu^*, C)$ decreases as k increases
⇒ can't use F to pick k
- other tech things related to cross validation or regularization

- K-means is a coordinate descent algorithm!

©Carlos Guestrin 2005-2014

14

Mixtures of Gaussians

Machine Learning – CSE546

Carlos Guestrin

University of Washington

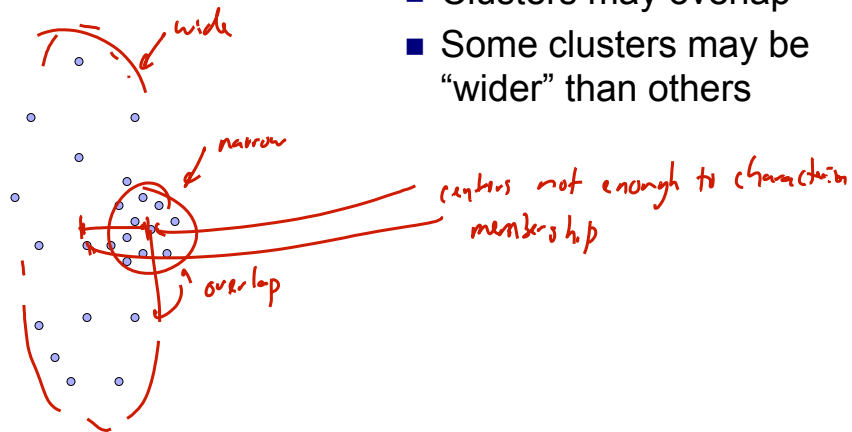
November 4, 2014

©Carlos Guestrin 2005-2014

15

(One) bad case for k-means

- Clusters may overlap
- Some clusters may be “wider” than others



©Carlos Guestrin 2005-2014

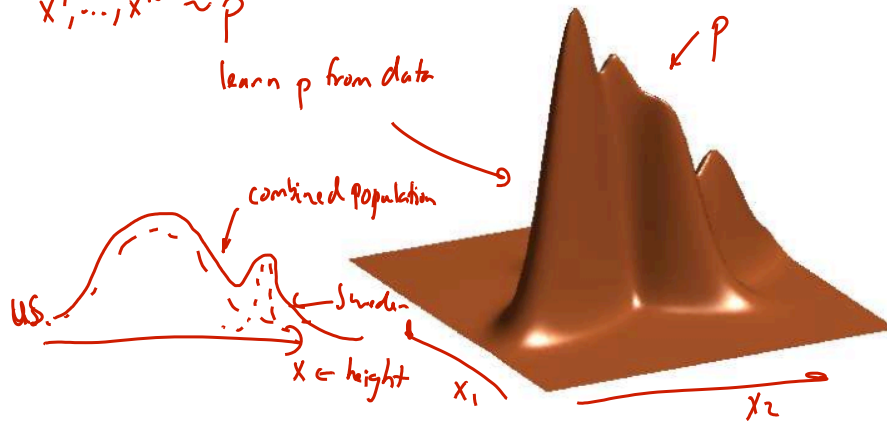
16

Density Estimation

- Estimate a density based on x^1, \dots, x^N

$$x^1, \dots, x^N \sim p$$

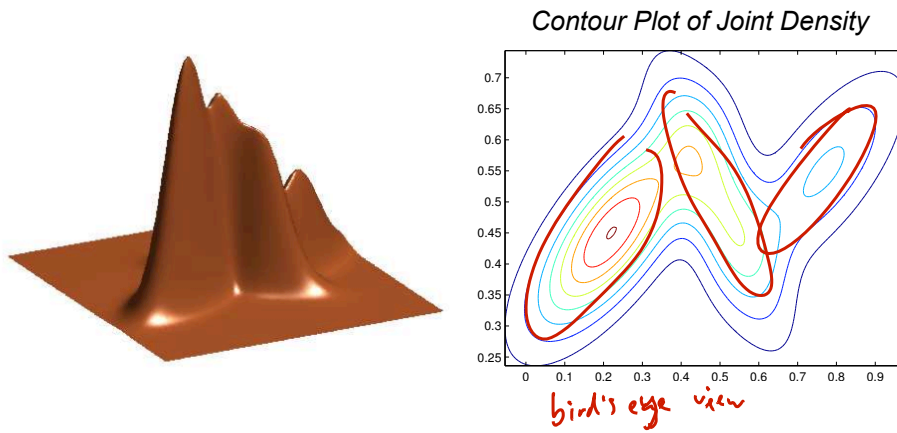
learn p from data



©Carlos Guestrin 2005-2014

17

Density Estimation



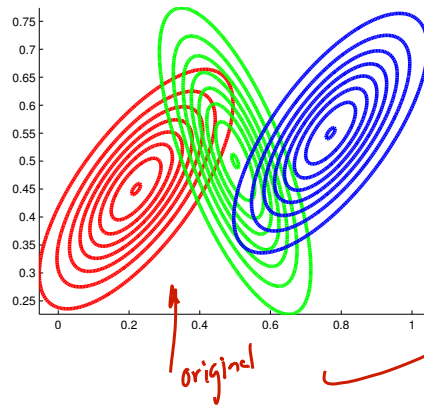
©Carlos Guestrin 2005-2014

18

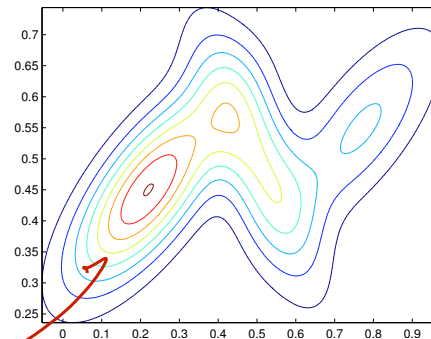
Density as Mixture of Gaussians

- Approximate density with a mixture of Gaussians

Mixture of 3 Gaussians



Contour Plot of Joint Density



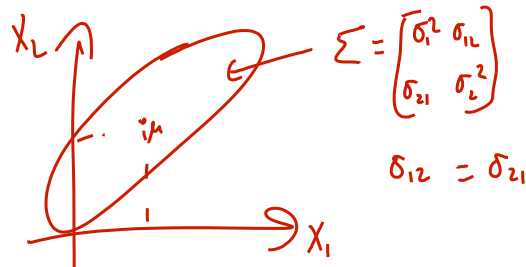
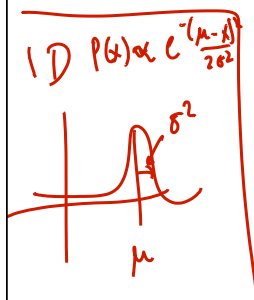
©Carlos Guestrin 2005-2014

19

Gaussians in d Dimensions

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right]$$

mean vector
covariance matrix



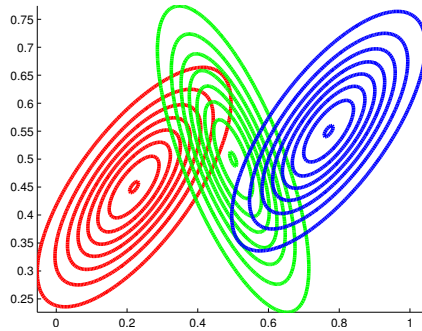
©Carlos Guestrin 2005-2014

20

Density as Mixture of Gaussians

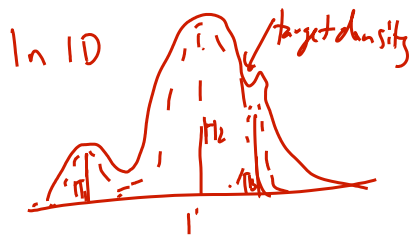
- Approximate density with a mixture of Gaussians $\pi_1, \pi_2, \dots, \pi_k$

Mixture of 3 Gaussians



$$p(x^i | \pi, \mu, \Sigma) = \sum_{i=1}^k \pi_i N(x^i | \mu_i, \Sigma_i)$$

Handwritten notes: $\pi_1, \pi_2, \dots, \pi_k$ are weights, $\sum_{i=1}^k \pi_i = 1$.



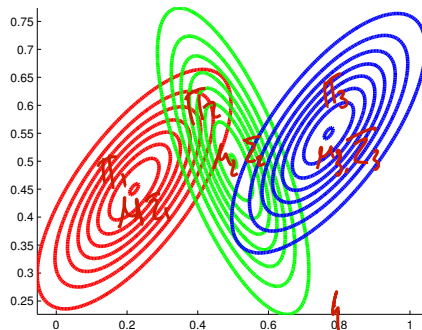
©Carlos Guestrin 2005-2014

21

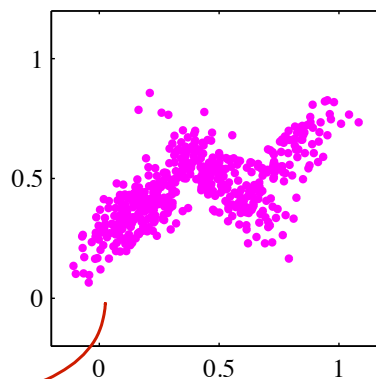
Density as Mixture of Gaussians

- Approximate with density with a mixture of Gaussians

Mixture of 3 Gaussians



Our actual observations



Handwritten notes: recover original densities How??

C. Bishop, Pattern Recognition & Machine Learning