Unsupervised learning, Clustering

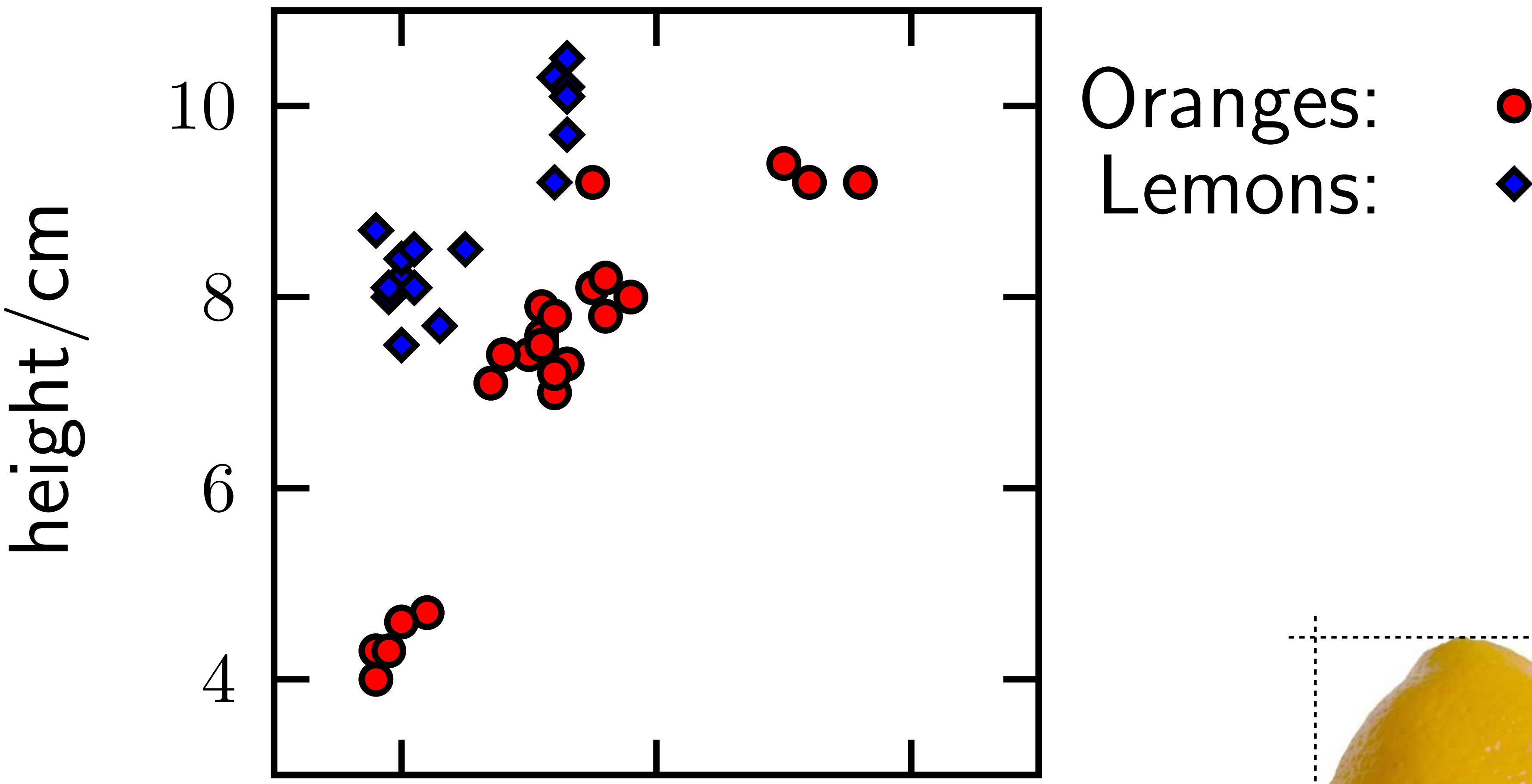
“Human brains are good at finding regularities in data.

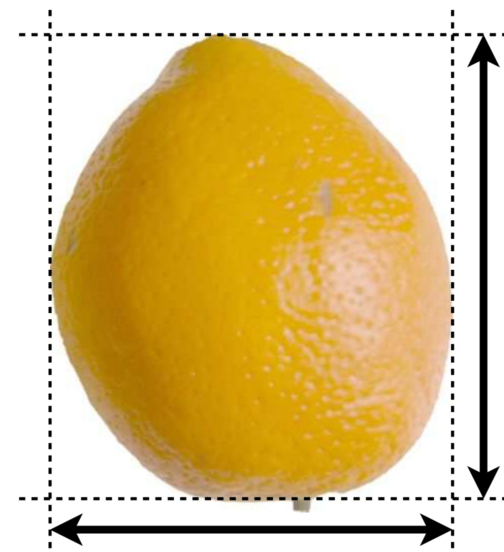
One way of expressing regularity is to put a set of objects into groups that are similar to each other.

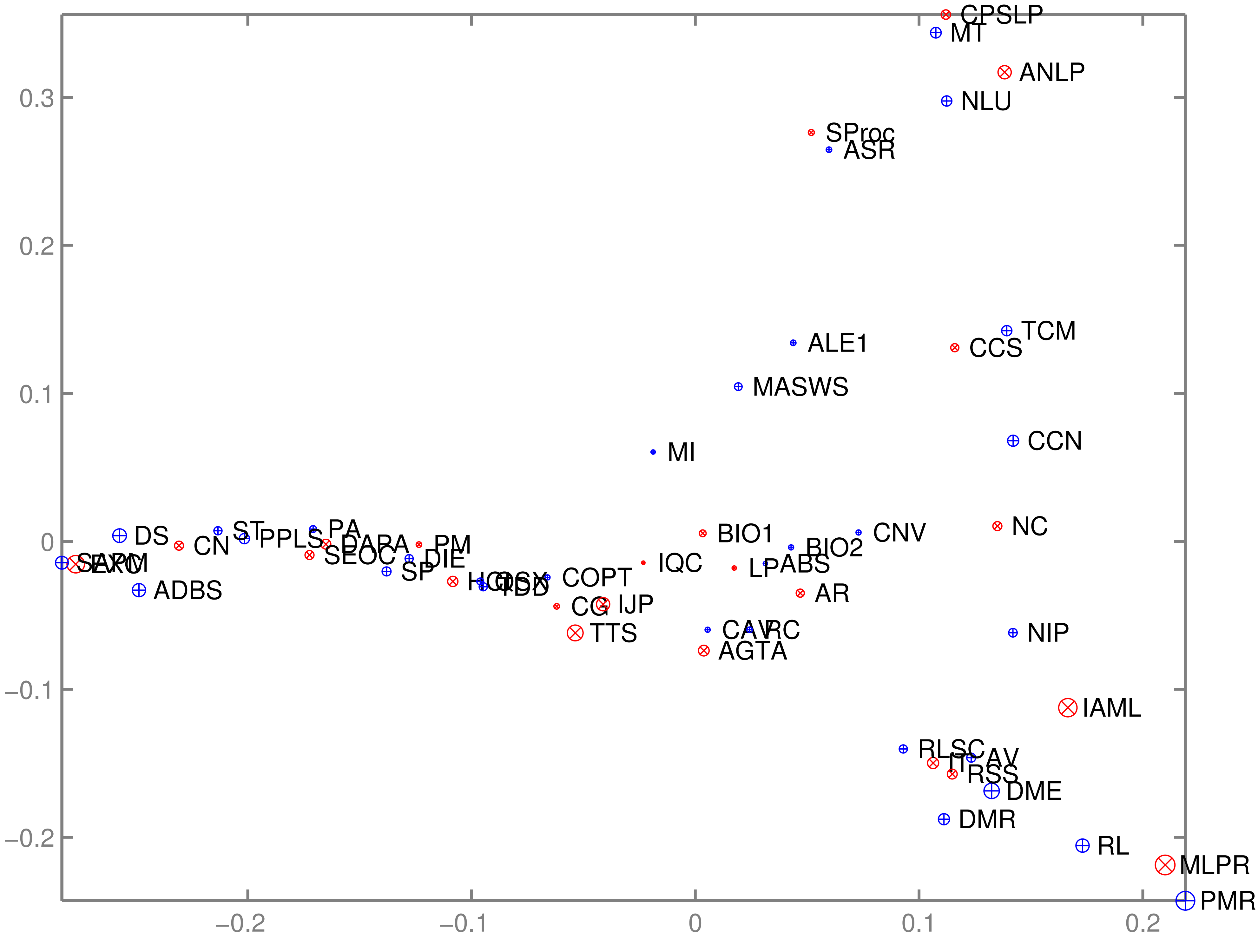
For example, biologists have found that most objects in the natural world fall into one of two categories: things that are brown and run away, and things that are green and don’t run away. The first group they call animals, and the second, plants.”

— David MacKay, ITILA textbook p284

Oranges and Lemons data







CPSLP
MT
ANLP
NLU
SProc
ASR
TCM
CCS
ALE1
MASWS
CCN
MI
DS
ST
PA
CN
PPLS
DAPA
PM
SEOC
DIE
SP
ADBS
COPT
CG
IJP
TTS
IQC
BIO1
BIO2
CNV
NC
AR
AGTA
NIP
IAML
RLSC
AV
RSS
DME
DMR
RL
MLPR
PMR
0.3
0.2
0.1
0
−0.1
−0.2
−0.2
−0.1
0
0.1
0.2

Stanley

Stanford Racing Team; DARPA 2005 challenge

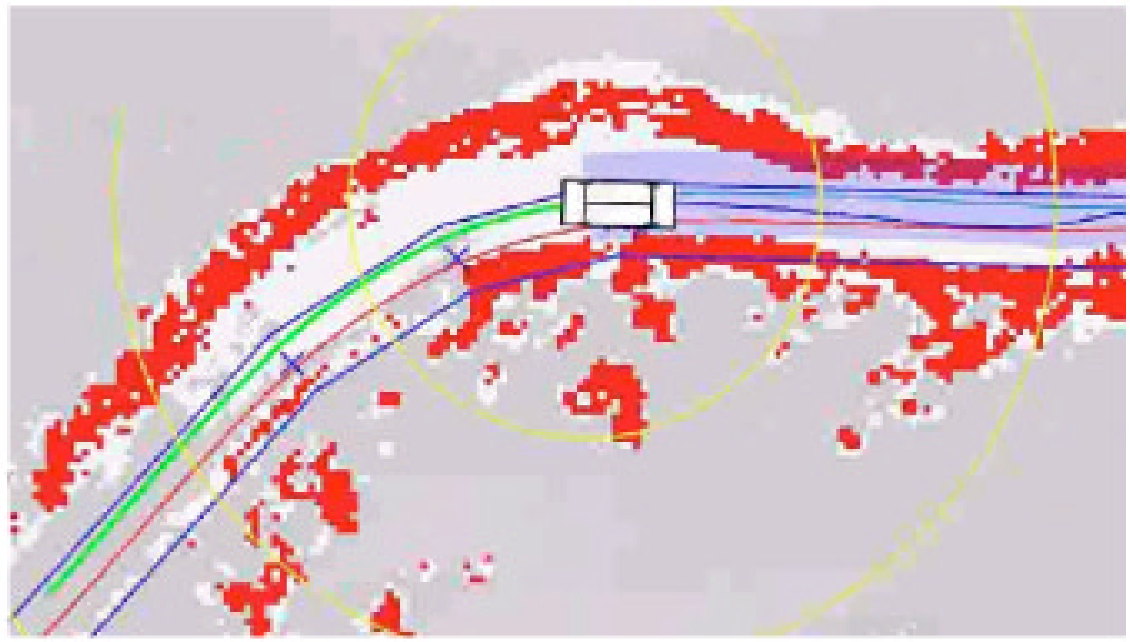
http://robots.stanford.edu/talks/stanley/

How to stay on a road?

Perception and intelligence

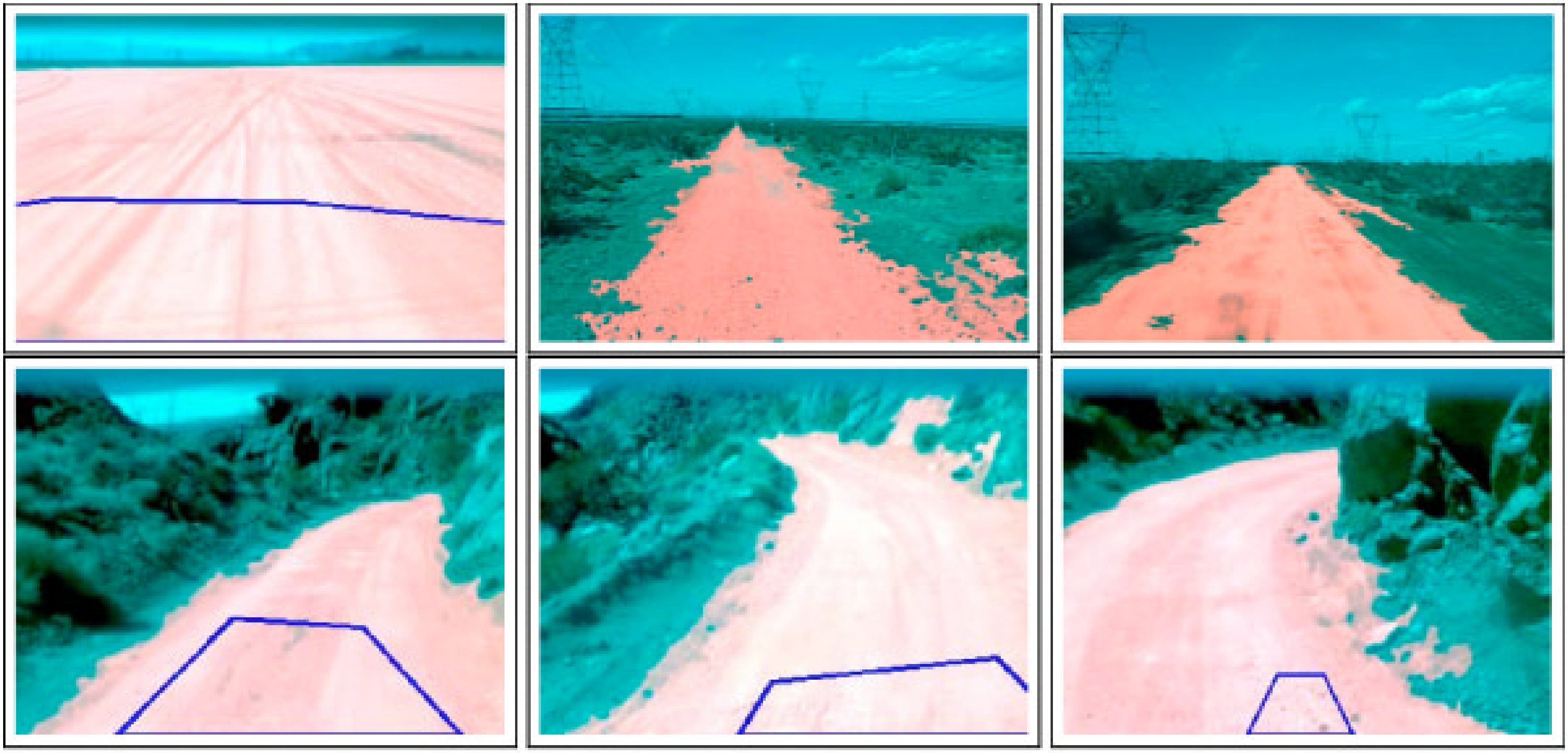
(a) **Beer Bottle Pass**

(b) **Map and GPS corridor**



It would look pretty stupid to run off the road,
just because the trip planner said so.

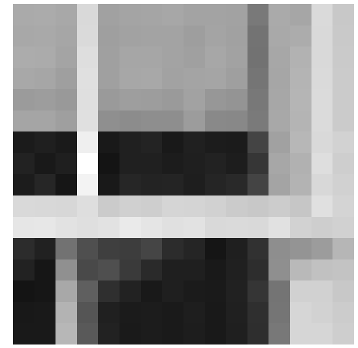
Clustering to stay on the road



Stanley used a Gaussian mixture model.
The cluster just in front is road (unless we already failed).

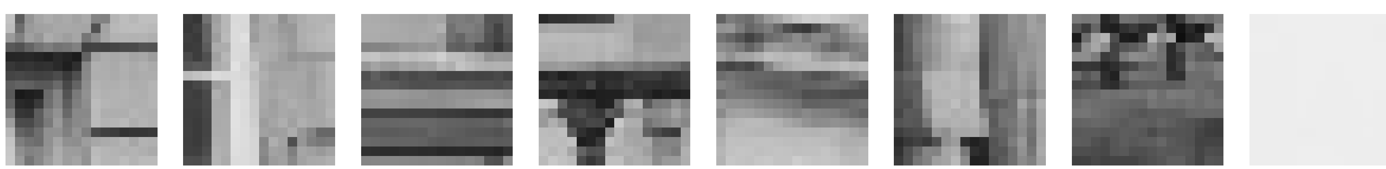
Example: Image denoising

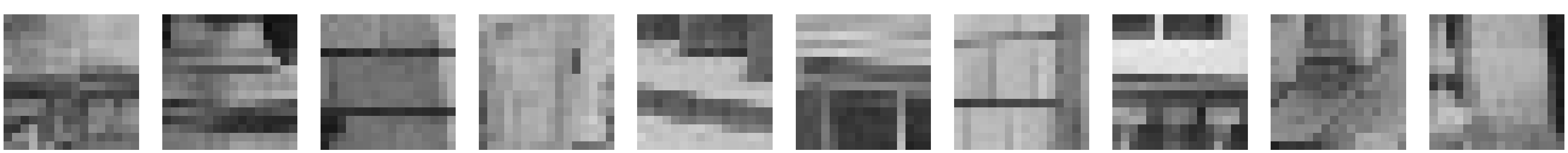
$\mathbf{y}$: 

$\mathbf{x}$: 

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x})$$

Likelihood: e.g. $\mathcal{N}(\mathbf{y};\, \mathbf{x},\, \sigma^2 I)$

Prior samples: 

 . . .

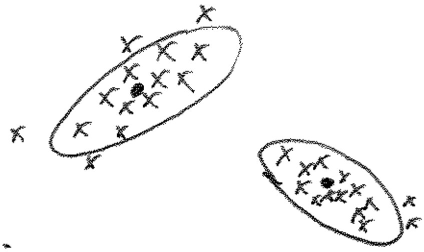
Zoran and Weiss, ICCV 2011

(a) Blurred (b) Krishnan et al. (c) EPLL GMM

$p(\mathbf{x})$ = Mixture of Gaussians fitted to patches

Mixtures of Gaussians

Model of a cluster



→ Model non-Gaussian distributions

Model

Hidden or latent variables:

$$z^{(n)} \sim \text{Discrete}(\underline{\pi})$$

(positive vector of length K that sums to one)

$$z^{(n)} \in \{1, 2, \ldots, K\}$$

Observations

If $z^{(n)} = k$, $\underline{x}^{(n)} \sim \mathcal{N}(\underline{x}^{(n)}; \underline{\mu}^{(k)}, \Sigma^{(k)})$

Likelihood of the model Parameters

Params: $\theta = \{\underline{\pi}, \{\mu^{(k)}, \Sigma^{(k)}\}_{k=1}^{K}\}$

$$P(X \mid \theta) = \sum_{\underline{z}} P(X, \underline{z} \mid \theta)$$

$$= \sum_{\underline{z}} P(X \mid \underline{z}, \theta)\, P(\underline{z} \mid \theta)$$

(X ↑ or "D" — no z's!)

$$= \underbrace{\sum_{\underline{z}}}_{\to \sum_{z^n}} \prod_n N(\underline{x}^{(n)}; \mu^{(z^{(n)})}, \Sigma^{(z^{(n)})}) \pi_{z^{(n)}}$$

$$\log P(X|\theta) = \sum_n \log \Big[\sum_{z^n} \pi_{z^n} N(\underline{x}^{(n)}; \mu^{z^n}, \Sigma^{z^n}) \Big]$$

$z^n \in \{1, 2, \ldots K\}$

Gradient-based fitting

Initialize carefully, maybe set $\{\Sigma^{(k)}\}$ broad.... so all probs reasonable

$\underline{\pi}$ constrained, $\sum_k \pi_k = 1$, $\pi_k > 0$

$\Sigma^{(k)}$ constrained, +ve definite

To optimize $\underline{\pi}$ optimize some other vector $\underline{c}$

$$\underline{\pi} = \text{softmax}(\underline{c})$$

$$\pi_k = \frac{e^{c_k}}{\sum_j e^{c_j}}$$

$\tilde{L}$ arbitrary matrix (lower-triangular)

$\downarrow$

$$L = \begin{cases} L_{ij} = e^{\tilde{L}_{ii}} & i = j \\ L_{ij} = \tilde{L}_{ij} & i > j \\ L_{ij} = 0 & i < j \end{cases}$$

$\downarrow$

$$\Sigma = LL^T$$

$\downarrow$

log-Likelihood

The alternative EM

Idea:

Pretend we observe $\{z^{(n)}\}$

Responsibility initialize:

$$r_k^{(n)} = \begin{cases} 1 & \text{if } z^{(n)} = k \\ 0 & \text{otherwise} \end{cases}$$

Maximize likelihood parameters

$$\pi_k = \frac{r_k}{N}, \qquad r_k = \sum_{n=1}^{N} r_k^{(n)}$$

$$\underline{\mu}^{(k)} = \frac{1}{r_k} \sum_{n=1}^{N} r_k^{(n)} \underline{x}^{(n)}$$

$$\Sigma^{(k)} = \frac{1}{r_k} \sum_{n=1}^{N} r_k^{(n)} \underline{x}^{(n)} \underline{x}^{(n)T} - \underline{\mu}^{(k)} \underline{\mu}^{(k)T}$$

EM Algorithm

0) Initialize params Θ.

1) <u>E-step</u>

$$r_k^{(n)} = P(z^{(n)} = k \mid \underline{x}, \Theta)$$

2) <u>M-step</u>

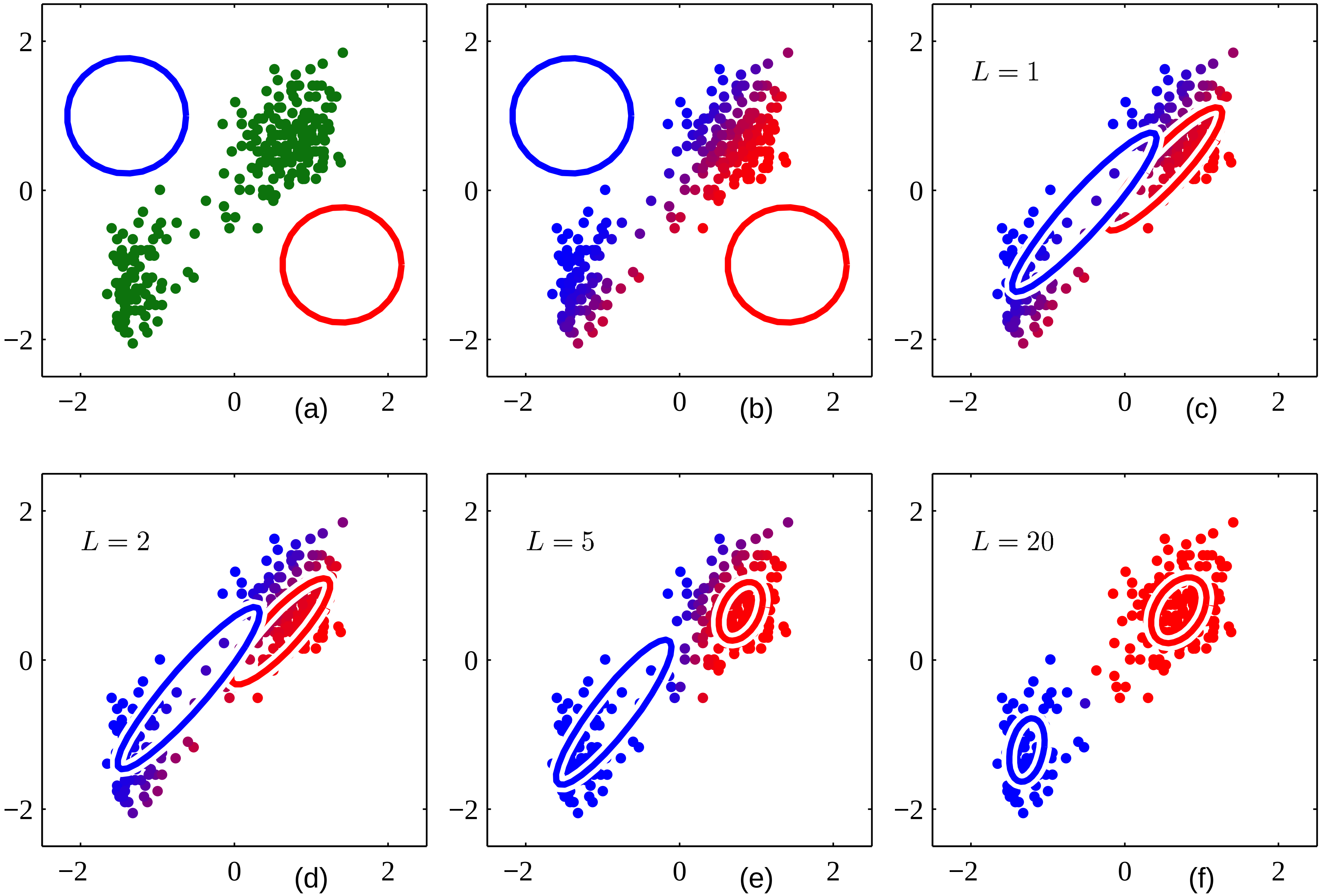
Use these real-valued $r_k^{(n)}$

In those equations

to fit Θ

3) GoTo 1) If not converged.

EM algorithm for Gaussian mixtures



Bishop Figure 9.8, or see Murphy p353