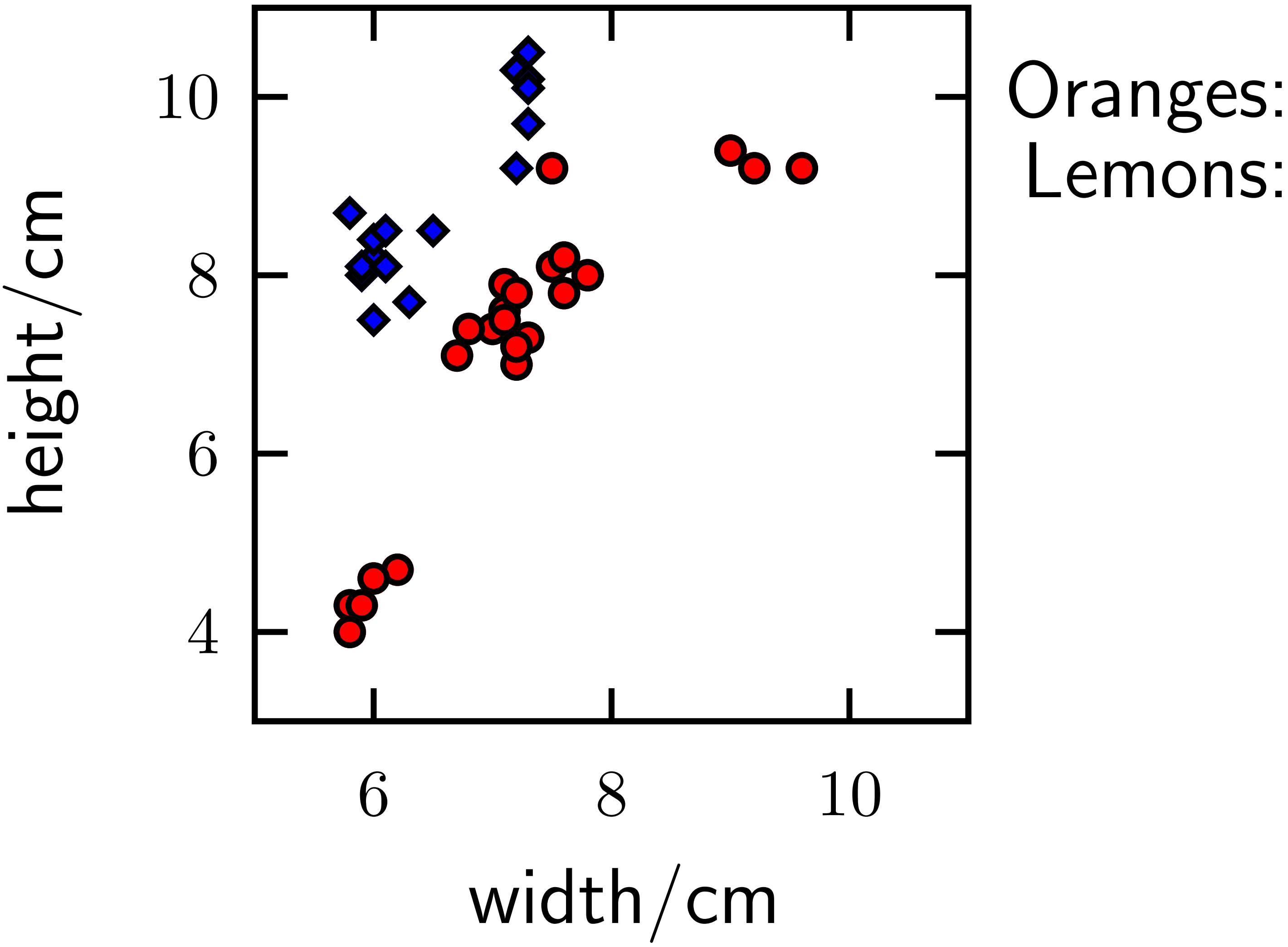
Clustering

"Human brains are good at finding regularities in data.

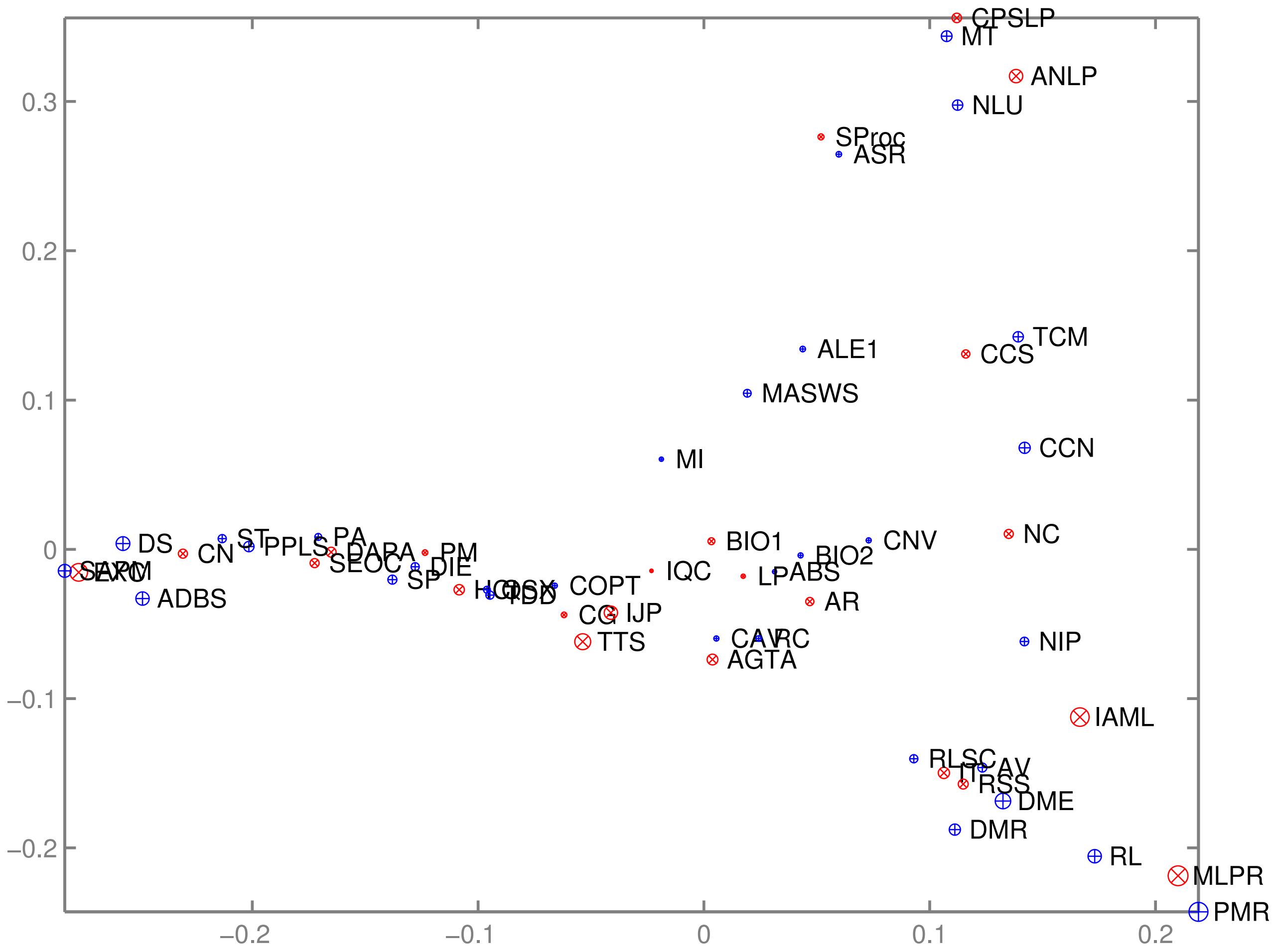
One way of expressing regularity is to put a set of objects into groups that are similar to each other.

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:
Lemons:
height/cm
width/cm
10
8
6
4
6
8
10



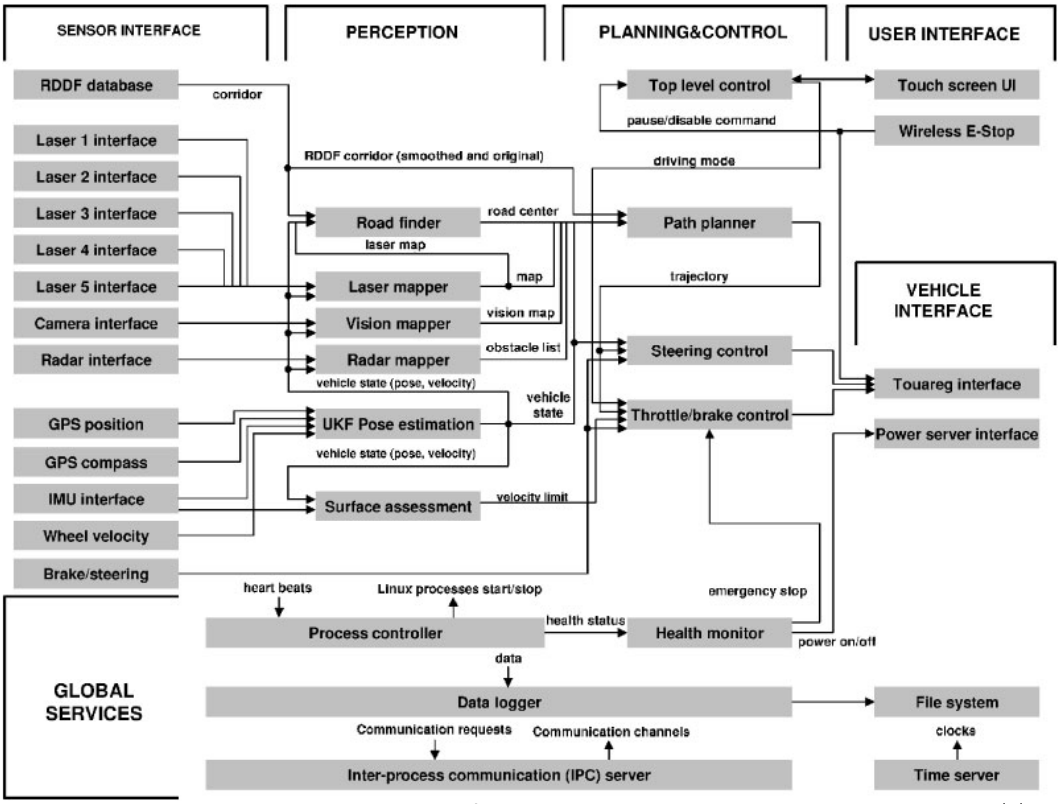
CPSLP
MT
ANLP
NLU
SProc
ASR
ALE1
TCM
CCS
MASWS
CCN
MI
NC
DS
CN
ST
PPLS
PA
DAPA
PM
SEOC
DIE
SP
ADBS
COPT
IQC
BIO1
BIO2
LPABS
CNV
AR
CG
IJP
TTS
AGTA
NIP
IAML
RLSC
AV
RSS
DME
DMR
RL
MLPR
PMR
0.3
0.2
0.1
0
−0.1
−0.2
−0.2
−0.1
0
0.1
0.2

Stanley

Stanford Raing Team; DARPA 2005 challenge

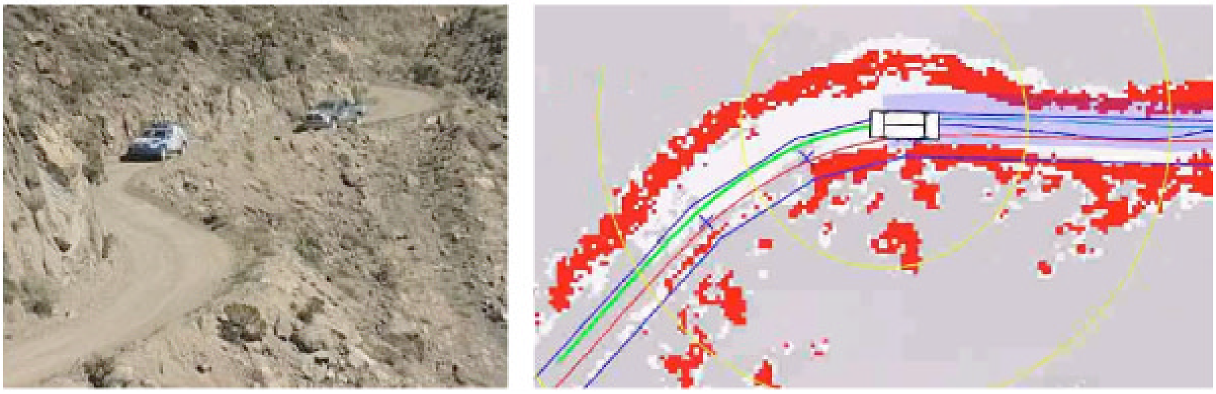
`http://robots.stanford.edu/talks/stanley/`

Inside Stanley



Stanley figures from Thrun et al., J. Field Robotics 23(9):661, 2006.

Perception and intelligence



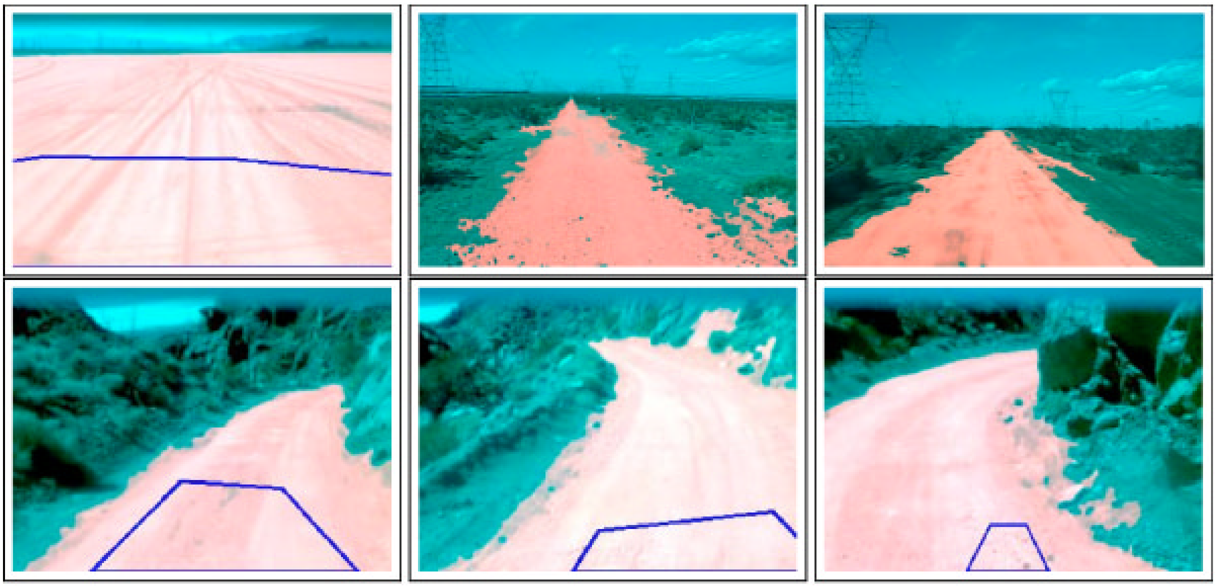
(a) Beer Bottle Pass (b) Map and GPS corridor

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

How to stay on the road?

Classifying road seems hard. Colours and textures change: road appearance in one place may match ditches elsewhere.

Clustering to stay on the road



Stanley used a Gaussian mixture model. "Souped up K-means."
The cluster just in front is road (unless we already failed).

Another example of learning $\underline{x}$'s

Patch of an image $\underline{x}$

I observe a noisy image $y \sim p(y|\underline{x})$

eg. $\mathcal{N}(\underline{y}; \underline{x}, \sigma_n^2 I)$

or convolution (a blur) + noise.

What's the underlying image?

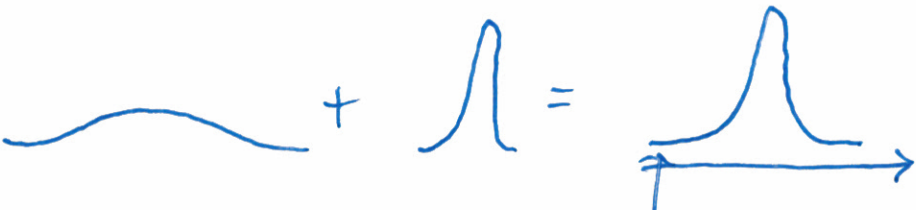
$$p(\underline{x}|y) \propto p(y|\underline{x})\, p(\underline{x})$$

One approach: optimize to find most probable image.

Mixtures of Gaussians

Model clusters

Model non-Gaussian distributions



Heavy tails.

Model

Hidden or latent variables:

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

positive vector of length K that sums to one

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

Observations

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

Likelihood of the model parameters

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

$$P(D \mid \theta) = \sum_{\underline{z}} P(D, \underline{z} \mid \theta) \quad \text{(sum rule)}$$

$\{ \underline{x}^{(n)} \}$
not including z!

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

$$= \sum_{\underline{z}} \prod_n p(\underline{x}^{(n)} \mid \underline{z}^{(n)}, \theta)\, p(z^{(n)} \mid \theta)$$

$$= \prod_n \sum_{z_n} \underbrace{p(\underline{x}^{(n)} \mid \underline{z}^{(n)}, \theta)}_{N(\underline{x}^{(n)};\, \mu^{(z^{(n)})}, \Sigma^{(z^{(n)})})} \overbrace{p(z^{(n)} \mid \theta)}^{\pi_{z^{(n)}}}$$

$$\log P(D \mid \theta) = \sum_{n=1}^{N} \log \left[\sum_k \pi_k\, N(\underline{x}^{(n)}; \mu^{(k)}, \Sigma^{(k)}) \right]$$

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

$\Sigma^{(k)}$ must be positive definite

Can do gradient fitting

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

$$\downarrow$$

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

$$\downarrow$$

$$\Sigma = L L^T$$

We optimize $\tilde{L}$

Model gives us log-likelihood & gradients of Σ

Backpropagate those derivates to $\tilde{L}$

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

$$\pi_i = \frac{e^{c_i}}{\sum_j e^{c_j}}$$

The alternative: EM

Idea:

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

Responsibility

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

Maximum Likelihood params:

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

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

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

"$\mathbb{E}[x x^T] - \mathbb{E}[x]\mathbb{E}[x^T]$"

EM Algorithm:

0) Initialize params Θ, set means different (or covs)

1) E-step

Soft responsibilities

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

$$= \frac{\pi_k N(\underline{x}^{(n)}; \underline{\mu}^{(k)}, \Sigma^{(k)})}{\sum_{k'} \pi_{k'} N(\underline{x}^{(n)}; \underline{\mu}^{(k')}, \Sigma^{(k')}}$$

2) M-step

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

In max likelihood fit eq's on prev page.

To fit Θ

3) Goto 1) If not converged.

Idea behind theory

Bound-based optimizer