“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

<table>
<thead>
<tr>
<th>height/cm</th>
<th>width/cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Oranges:  
Lemons:
Stanley

Stanford Racing Team; DARPA 2005 challenge

http://robots.stanford.edu/talks/stanley/
How to stay on a road?
Perception and intelligence

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

\[ p(x | y) \propto p(y | x) p(x) \]

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

Prior samples:
<table>
<thead>
<tr>
<th>Kernel 1</th>
<th>Kernel 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 × 17</td>
<td>19 × 19</td>
</tr>
<tr>
<td>25.84</td>
<td>27.17</td>
</tr>
</tbody>
</table>


**Figure 8: Deblurring experiments**

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

\[ p(x) = \text{Mixture of Gaussians fitted to patches} \]
Variational Inference

\[ \text{Cost function:} \quad J = -E_q [ \log p(D|w)] - E_q [\log p(w)] + E_q [\log q(w)] \]

\[ = E [\text{neg. log likelihood}] - H[q] + D_{KL}(q(w) \parallel p(w)) \]

Marginal Likelihood bound

\[ \log p(D) \geq -J \]

Minimize \( J \) wrt \( (m, V) \) and \( \sigma_w, \sigma_y, \ldots \)

Approx. posterior well

Find good model
Minimize $J$

Like SGD, but need some tricks

**Trick #1 Unconstrained Optimization**

If we do SGD on $6_y^2$ might get re values

Optimize $\log 6_y^2$ instead

$V$ positive definite, symmetric

$V = LL^T$, $L$ lower triangular

Diagonal is $+$ve.

We create another matrix

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

Optimizer has $\tilde{L} \rightarrow L \rightarrow V = LL^T \rightarrow \ldots$ cost

exp diag

$SGD$ on $\tilde{L} \leftarrow$ backprop
Evaluating the cost

\[ D_{KL}(q \| p) \], or "Entropy terms"

We can evaluate these.

Likelihood term:

\[
\mathbb{E}_q \left[ \log p(D(w)) \right] \\
= \mathbb{E}_q \left[ \sum_{n=1}^{N} \log p(y^{(n)} | x^{(n)}, w) \right]
\]

For logistic regression: \( \to \) 1D integral \( \to \) do numerically.

Stochastic estimate: Trick #2 Reparameterization

\[
\mathbb{E}_{N(w; m, \nu)} \left[ f(w) \right] \\
= \mathbb{E}_{N(z; 0, I)} \left[ f(m + Lz) \right]
\]

Sample \( w \), by \( z \sim N(0, I) \)

\[ w = m + Lz \]
Monte Carlo estimate

\[ x = \frac{1}{S} \sum_{s=1}^{S} f(m + L \xi^{(s)}) \]

\[ S = 1, \quad \xi^{(s)} \sim N(0, I) \]

\[ x \sim f(m + L \xi), \quad \xi \sim N(0, I) \]

\[ \nabla_m E_N(w; m, \nu) [f(w)] \]

\[ \sim \nabla_m f(m + L \xi) \]

\[ \nabla_L E_N(w; m, \nu) [f(w)] \]

\[ \sim \nabla_L f(m + L \xi) \]

\[ \nabla_w f(w) \bigg|_{w = m + L \xi} \to^T \]

Need derivatives of log likelihood as usual.

(Also find \( \nabla \xi \) for SGD from \( \nabla L \))
Mixtures of Gaussians

Road.