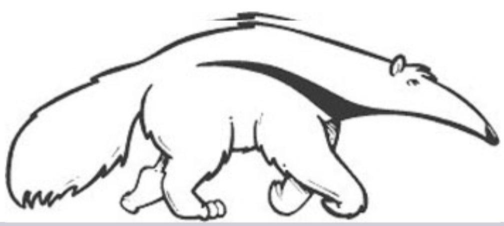
Maximum Entropy Learning

Learning in Graphical Models

Prof. Alexander Ihler



BREN:ICS
INFORMATION AND COMPUTER SCIENCES

UNIVERSITY of CALIFORNIA IRVINE

Maximum Entropy Learning

$$\max_{q(x)} \mathbb{H}(q) \qquad s.t.\ \forall i \quad \mathbb{E}_q[u_i(x)] = \hat{\mu}_i$$

Shorthand:

$$\mu_i(q) = \mathbb{E}_q[u_i(x)]$$

- Maximum Entropy Learning Principle
  - Identify some features $u_i(x)$, and their desired expectations
  - Find the most “agnostic” distribution that matches the data
  - q(x) can be any distribution, of any form

Maximum Entropy Learning

$$\max_{q(x)} \mathbb{H}(q) \qquad s.t. \ \forall i \ \ \mathbb{E}_q[u_i(x)] = \hat{\mu}_i$$

Shorthand:

$$\mu_i(q) = \mathbb{E}_q[u_i(x)]$$

Lagrangian:

$$\max_{q(x)} \min_{\theta} \ \mathbb{H}(q) + \theta \cdot (\mu(q) - \hat{\mu})$$

Define:

$$f(x) = \exp(\theta \cdot u(x))$$

$$Z = \textstyle\sum_x f(x)$$

$$p(x) = Z^{-1} f(x)$$

then,

$$D(q||p)$$

$$= \mathbb{E}_q[\log q - \log p]$$

$$= -\mathbb{H}(q) - \theta \cdot \mu(q) + A(\theta)$$

Dual: $\min_{\theta} \ \max_{q(x)} \ \mathbb{H}(q) + \theta \cdot (\mu(q) - \hat{\mu})$

$$\Leftrightarrow \min_{\theta} \ \max_{q(x)} \ -D(q||p) + A(\theta) - \theta \cdot \hat{\mu}$$

$$\Rightarrow q^*(x) = p(x) \quad \Rightarrow \quad \min_{\theta} A(\theta) - \theta \cdot \hat{\mu}$$

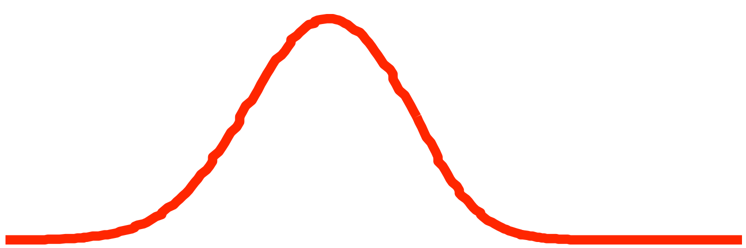
Optimal q* is exponential family with features u(x)!

Optimal parameters θ are the MLE estimates!

Maximum Entropy Learning

- Example: Gaussian distribution
  - Exponential family with features x, x^2
  - Has the highest entropy of any distribution with variance *v*

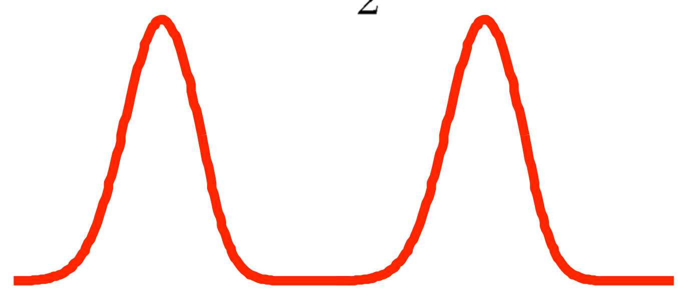
$$\mathcal{N}(x\,;\,0,1)$$



$$\mathbb{E}[(x-\mu)^2] = 1$$

$$\mathbb{H}[x] \approx 1.41$$

$$\frac{1}{2}\mathcal{N}(x\,;\,-1,.25) + \frac{1}{2}\mathcal{N}(x\,;\,1,.25)$$



$$\mathbb{E}[(x-\mu)^2] \approx 1$$

$$\mathbb{H}[x] \approx 0.73$$