Probability Density Estimation

Machine Learning - 10601

Geoff Gordon, Miroslav Dudík
http://www.cs.cmu.edu/~ggordon/10601/
September 9, 2009

Machine learning challenges

Representation of

- observations, assumptions, solutions

Generalization to

- present (but unobserved) data, future data

Computation

Last few lectures

- how to describe probability distributions

REPRESENTATION

Next

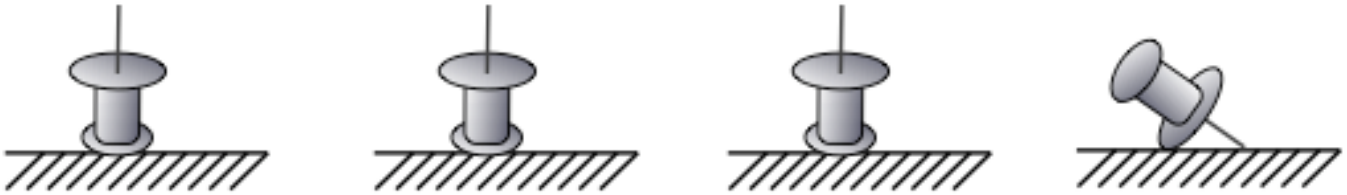
- how to learn probability distributions

GENERALIZATION

Probability Density Estimation

thumbtack
- up “heads” w/prob θ
- down “tails” w/prob $1-\theta$

unknown probability $0 \leq \theta \leq 1$



θ = ?

Maximum Likelihood Estimation

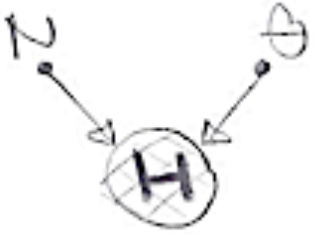
- a priori no reason to favor some θ over another
- pick the one that assigns the largest probability to data

EXPERIMENT:

N tosses

H came up heads

N-H came up tails



$$p(H \mid N, \theta) = \binom{N}{H} \theta^{H} (1-\theta)^{N-H}$$

binomial

$$\max_{\theta} p(\underline{\underline{H}} \mid N, \underline{\underline{\theta}})$$

data

unknown, but fixed parameter

instead: $\max_{\theta} \left(\log p(H \mid N, \theta) \right)$

$$\log p(H \mid N, \theta) = \text{const.} + H \log \theta + (N-H) \log(1-\theta)$$

as a function of θ

$$\frac{\partial \log p(H \mid N, \theta)}{\partial \theta} = \frac{H}{\theta} - \frac{N-H}{1-\theta} = 0$$

$$(1-\theta) H = (N-H)\theta$$

$$H = N\theta$$

$$\theta = \frac{H}{N}$$

empirical frequency