Expectation Maximization

CS 536: Machine Learning

Littman (Wu, TA)

Administration

Have you seen HMMs?

Reviews:

- http://www.cs.rutgers.edu/~mlittman/courses/ml03/iCML03/review.html

Any question about these?

No discussion period.

- 24 Nov 2003 --- Reviews due, returned
  - Improve papers, work on slides
- 03 Dec 2003 --- Oral decision announced
- 08 Dec 2003 --- Workshops (2 or 3 rooms)
- 10 Dec 2003 --- Talks, final papers due (pdf)

Some Reading

Learning from Data

We want to learn a model with a set of parameter values M.

We are given a set of data D.

An approach: $\text{argmax}_M \Pr(M|D)$

This is the *maximum likelihood* model (ML).

How relate to $\Pr(D|M)$?

Super Simple Example

Coin I and Coin II. (Weighted.)

Pick a coin at random (uniform).

Flip it 4 times.

Repeat.

What are the parameters of the model?

Data

Coin I	Coin II
HHHT	TTTH
HTHH	THTT
HTTH	TTHT
THHH	HTHT
HHHH	HTTT

Probability of *D* Given *M*

p: Probability of H from Coin I

q: Probability of H from Coin II

Let's say *h* heads and *t* tails for Coin I.
h' and *t*' for Coin II.

$\Pr(D|M) = p^h (1-p)^t q^{h'} (1-q)^{t'}$

How maximize this quantity?

Maximizing *p*

$D_p(p^h (1-p)^t q^{h'} (1-q)^{t'}) = 0$

$D_p(p^h)(1-p)^t + p^h D_p((1-p)^t) = 0$

$h\ p^{h-1} (1-p)^t = p^h\ t(1-p)^{t-1}$

$h (1-p) = p\ t$

$h = p\ t + h\ p$

$h/(t+h) = p$

Missing Data

HHHT	HTTH
TTTH	HTHH
THTT	HTTT
TTHT	HHHH
THHH	HTHT

Oh Boy, Now What!

If we knew the labels (which flips from which coin), we could find ML values for p and q.

What could we use to label?

- p and q!

Computing Labels

$p = 3/4$, $q = 3/10$
Pr(Coin I | HHTH)
= Pr(HHTH | Coin I) Pr(Coin I) / c
= $(3/4)^3 (1/4) (1/2)/c = .05273438/c$
Pr(Coin II | HHTH)
= Pr(HHTH | Coin II) Pr(Coin II) / c
= $(3/10)^3 (7/10) (1/2)/c = .00945/c$
c = .05273438 + .00945,
so 84.8%/15.2%

Expected Labels

	I	II		I	II
HHHT	.85	.15	HTTH	.44	.56
TTTH	.10	.90	HTHH	.85	.15
THTT	.10	.90	HTTT	.10	.90
TTHT	.10	.90	HHHH	.98	.02
THHH	.85	.15	HTHT	.44	.56

Wait, I Have an Idea

Pick some model M_0

Expectation

- Compute expected labels via M_i

Maximization

- Compute ML model M_{i+1}

Repeat

Could This Work?

Expectation-Maximization (EM)

Pr($D|M_i$) will not decrease.

Another optimization search algorithm. (How else could we do this?)

Coin Example

Compute expected labels.

Compute counts of heads and tails (fractions).

Divide to get new probabilities.

p=.63 q=.42 Pr(D|M)=9.95 x 10^{-13}

p=.42 q=.63 Pr(D|M)=9.95 x 10^{-13}

p=.52 q=.52 Pr(D|M)=9.56 x 10^{-13}

More General EM

Need to be able to compute probabilities: generative model

Need to tabulate counts to estimate ML model

Let's think this through with mixtures of Gaussians or HMMs…