MACHINE INTELLIGENCE

UNIT - 4

Expectation Maximisation

feedback/corrections: vibha@pesu.pes.edu

VIBHA MASTI

Expectation Maximisation

- Iterative way to learn max likelihood estimate of parameters
- Latent variables
- Eg: HMM — learning problem of A and B
- Things we will learn
  - Binomial mixture model
  - Gaussian mixture model
  - k-means clustering

E-Step

- Estimate latent variables using the observed data and the current estimate of model parameters
- Initial estimate of model parameters is random

M-Step

- Maximise likelihood function of model params under the assumption that the missing data are known

1. Binomial Mixture Models

1.1 Simple case: 2 coins A & B

- To find: bias of each coin towards head (p_A and p_B)
- Mathematical terms: estimate parameter p of two binomial distributions
- If all values known,

coin	flips	# coin A heads	# coin B heads
B	HTTTHHTHTH	0	5
A	HHHHTHHHHH	9	0
A	HTHHHHHTHH	8	0
B	HTHTTTHHTT	0	4
A	THHHTHHHTH	7	0

$$p_A = \frac{24}{30} = 0.8 \qquad p_B = \frac{9}{20} = 0.45$$

- If latent variables present (identity of coin)

coin	flips	# coin A heads	# coin B heads
?	HTTTHHTHTH	?	?
?	HHHHTHHHHH	?	?
?	HTHHHHHTHH	?	?
?	HTHTTTHHTT	?	?
?	THHHTHHHTH	?	?

EM

- Let X represent a sequence of H & T
- Let Z_A, Z_B be the events of choosing A and B for one trial, respectively
- Let $P(Z_A) = P(Z_B) = 0.5$
- Let initial estimates of the bias $p_A = 0.6$ and $p_B = 0.5$
- Let $X =$ HTTTHHTHTH
- Conditional probability

$$P(X \mid Z_A) = (0.6)^5 (0.4)^5$$

$$P(X \mid Z_B) = (0.5)^5 (0.5)^5$$

- Bayes Theorem

$$P(Z_A \mid X) = \frac{P(X \mid Z_A)\, P(Z_A)}{P(X)} = 0.45$$

$$P(Z_B \mid X) = \frac{P(X \mid Z_B)\, P(Z_B)}{P(X)} = 0.55$$

- For n iterations

- Fill table with # heads and tails as $P(X|Z_i) \times$ # heads

i	flips	probability it was coin A $P(Z_{iA}|X_i)$	probability it was coin B $P(Z_{iB}|X_i)$	# heads attributed to A	# heads attributed to B
1	HTTTHHTHTH	0.45	0.55	2.2	2.8
2	HHHHTHHHHH	0.8	0.2	7.2	1.8
3	HTHHHHHTHH	0.73	0.27	5.9	2.1
4	HTHTTTHHTT	0.35	0.65	1.4	2.6
5	THHHTHHHTH	0.65	0.35	4.5	2.5

$$P_A = \frac{2.2+7.2+5.9+1.4+4.5}{10(0.45+0.8+0.73+0.35+0.65)} = 0.71$$

heads by coin B

$$P_B = \frac{2.8+1.8+2.1+2.6+2.5}{10(0.55+0.2+0.27+0.65+0.35)} = 0.58$$

total coin tosses by coin B

Use P_A, P_B to recompute $P(X|Z)$ and $P(Z|X)$

1.2 Generalise for k coins

- Assume k coins with prior distribution

$$\pi = \{\pi_1, \pi_2, \ldots, \pi_k\}$$

prob of picking a coin

$$\sum_{i=1}^{k} \pi_i = 1 \qquad 0 \le \pi_i \le 1$$

- If equally probable, $\pi_i = \frac{1}{k}$

- Define success probability (heads)

$$P = \{p_1, p_2, \dots, p_k\}$$

p_i = prob of heads with ith coin

- Introduce hidden bool Z_{ij}
  - instance i generated from coin j

$i = 1$ to n (no. of instances)

$j = 1$ to k (no. of coins)

- Z vector for every instance i

$$Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{ik})$$

where exactly one $Z_{ij} = 1$

i	x_i	z_{i1}	z_{i2}	z_{i3}	# heads
1	H T H H	0	1	0	3
2	H H T T	1	0	0	2
3	H H H H	0	1	0	4
4	H T H H	0	0	1	3
5	T T T H	0	0	1	1

$n = 5$
$m = 4$
$k = 3$

$p = \{p_1, p_2, p_3\}$

$$P(x_i, z_i \mid p) = \prod_{j=1}^{m} \prod_{l=1}^{k} \left(p_l^{x_{ij}} (1 - p_l)^{1 - x_{ij}} \right)^{z_{il}}$$

$$\hat{p}_1 = \frac{\sum_{i=1}^{n} z_{i1} \sum_{j=1}^{m} x_{ij}}{m * \sum_{i=1}^{n} z_{i1}}$$

2. Gaussian Mixture Models

- PDF of univariate ND

$$f(X = x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$X \sim N(\mu, \sigma^2)$$

- PDF of multivariate ND

$$x = [X_1, x_2, \dots, x_n]^T$$

$$f(X = x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

GMM

$$p(X = x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(X = x; \mu_k, \Sigma_k)$$

- Weighted mix of K Gaussians

$$0 \leq \pi_k \leq 1, \quad \sum_{i=1}^{K} \pi_k = 1$$

- Soft clustering

Q: Assume $K=2$, $\sigma_1^2 = \sigma_2^2 = \frac{1}{2}$

$x = (2, 4, 7)$

$\pi = (0.5, 0.5)$

$\mu_1 = 3 \quad \mu_2 = 6$

$a \sim N(3, 1/2)$

$b \sim N(6, 1/2)$

Probability that point x_i generated from G_1 (a) and G_2 (b)

calc z-score & find p

$$G_{11} = a_1 = P(a|x_1) = \frac{P(x_1|a)\,P(a)}{P(x_1|a)P(a) + P(x_1|b)P(b)}$$

$$a_1 = \frac{0.2076 \times 0.5}{0.2076 \times 0.5 + 6.35 \times 10^{-8} \times 0.5}$$

$$a_1 = 1$$

$$G_{12} = a_2 = P(a|x_2) = \frac{P(x_2|a)\,P(a)}{P(x_2|a)P(a) + P(x_2|b)P(b)}$$

$$a_2 = \frac{0.2076 \times 0.5}{0.2076 \times 0.5 + 0.0103 \times 0.5}$$

$$a_2 = 0.953$$

$$G_{13} = a_3 = P(a|x_3) = \frac{P(x_3|a)\,P(a)}{P(x_3|a)P(a) + P(x_3|b)P(b)}$$

$a_3 = 0$

$G_{21} = b_1 = P(b|x_2) = 0$

$G_{22} = b_2 = 0.047$

$G_{23} = b_3 = 1$

M step

$$\mu_1 = \frac{1 \times 2 + 0.953 \times 4 + 0 \times 7}{1 + 0.953 + 0} = 2.976$$

$$\mu_2 = \frac{0 \times 2 + 0.047 \times 4 + 1 \times 7}{0 + 0.047 + 1} = 6.865$$

can estimate σ^2 also