Notes of Expectation Maximization (EM) Algorithm

Zengchang Qin (PhD)

3 – Coin problem with Expectation Maximization

1) Given a coin with the probability p of being head.
The probability distribution is a Bernoulli distribution:

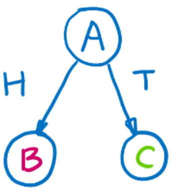
$$P(x) = p^x (1-p)^{1-x}$$

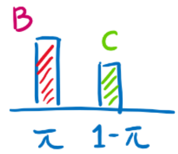
Given a sequence of results "HTTHH ⋯ H"
P is estimated by:

$$p = \frac{\sum_{i=1}^{N} I(x_i = H)}{N}$$, where $I(\cdot)$ is an indicator function.

2) There are three coins A, B and C with parameters π, p and q, respectively.

A $\sim \pi$
B $\sim p$
C $\sim q$





If the tossing result of A is H, we toss coin B, otherwise we toss the coin C, the tossing results of B and C are recorded as a sequence.

E.g. H H T T H T H T T H

3) If we assume to know which coin generates the result.
For example: H H T T H T H T H H
H H T T H T H T H H

The results of the coin B is in red and coin C is in green.
If we have the colored results, we can estimate parameters by:

3) If we assume to know which coin generates the result.

For example: HHTTHTHTHH

HHTTHTHTHH

The results of the coin B is in red and coin C is in green.

If we have the colored results, we can estimate parameters by:

4) For the above 3-coin problem, the parameter $\theta = (\pi, P, q)$

$$L(\theta) = \arg\max_{\theta} P(x \mid \theta)$$

$$P(x \mid \theta) = \sum_z P(x, z \mid \theta)$$
$$= \sum_z P(x \mid z, \theta) P(z \mid \theta)$$

where z is a latent (hidden) variable, $z = \{0, 1\}$

$z = 1$ indicates the case of coin B and $z = 0$ is for coin C. Therefore:

$$P(x \mid \theta) = \sum_z P(x \mid z, \theta) P(z \mid \theta)$$
$$= \underbrace{P(x \mid z=1, \theta)}_{\text{Bernoulli distribution}} \underbrace{P(z=1 \mid \theta)}_{\text{Prior}} + \underbrace{P(x \mid z=0, \theta)}_{\text{Bernoulli distribution}} P(z=0 \mid \theta)$$

$$P(x \mid z=1, \theta) = p^x (1-p)^{1-x}$$
$$P(x \mid z=0, \theta) = q^x (1-q)^{1-x}$$
$$P(z=1 \mid \theta) = \pi, \quad P(z=0 \mid \theta) = 1 - \pi$$
$$P(x \mid \theta) = \underbrace{\pi p^x (1-p)^{1-x}}_{\text{Coin B}} + \underbrace{(1-\pi) q^x (1-q)^{1-x}}_{\text{Coin C}}$$

We use μ to represent the probability of x is tossing by B given current parameter $\theta = (\pi, P, q)$

$$\mu = \frac{\pi p^x (1-p)^{1-x}}{\pi p^x (1-p)^{1-x} + (1-\pi) q^x (1-q)^{1-x}} \qquad x \in \{0, 1\}$$

$X=1$ indicates H and $X=0$ indicates T.

Given a sequence of tossing results $X=\{x_1, x_2 \cdots x_N\}$

$$\mu_i = \frac{\pi p^{x_i}(1-p)^{1-x_i}}{\pi p^{x_i}(1-p)^{1-x_i} + (1-\pi) q^{x_i}(1-q)^{1-x_i}}$$

5) Given a sequence of tossing results, we now have a probability estimation of being B or C.

H H T T H T H T H H

It is like H or T cannot be 100% of being red or green, it has a "probabilistic degree" of being red or green.

Based on Section (3) we can estimate θ using current μ.

$$\pi = \frac{\sum_{i=1}^{N} \mu_i}{N}$$ percentage of being red of all N tosses

$$p = \frac{\sum_{i=1}^{N} \mu_i x_i}{\sum_{i=1}^{N} \mu_i}$$ percentage of red H among red

$$q = \frac{\sum_{i=1}^{N} (1-\mu_i) x_i}{\sum_{i=1}^{N} (1-\mu_i)}$$ percentage of green H among green

6) This can be summarized as E-step and M-step.

E-step: Calculating μ using θ

M-step: recalculate θ using μ from the last step.