

הסכימה הכללית של EM

המטרה: שערך $\hat{\theta}$

אתחול: θ_0

איטרציה עד התכנסות:

צעד E: מחשבים $q(\theta_0) = P(x|y; \theta_0)$

צעד M: $\theta(q) = \operatorname{argmax}_{\theta} \sum_x P(x|y; \theta) \cdot \log P(x, y; \theta)$

$\theta_0 \leftarrow \theta(q)$

נפתח את משוואות E, M למודל עירוב היסטוגרמות

קודם נפתח רישום של המשוואות הכלליות של צעדי EM עבור y שהוא סדרת תצפיות:

$$y = (y_1, \dots, y_N)$$

$$x = (x^1, \dots, x^N)$$

כאשר x^t הוא ערך X^t - המשתנה החבוי של y_t
כאשר התצפית היא סדרת תצפיות, אפשר לחזור על פיתוח הסכימה הכללית, עם $q(x^t)$ לכל איבר בסדרה
עבור נוסחת הנראות

$$\log P(y; \theta) = \sum_{t=1}^N \log P(y_t; \theta)$$

ואז הסכימה הכללית למשוואות צעד E והM תהיה:

צעד E: נחשב:

$$\forall_{x_i \in X} q_t(\theta_0) = P(x^t = x_i | y_t, \theta_0) = w_{t_i}$$

צעד M:

$$\theta(q) = \operatorname{argmax}_{\theta} Q(\theta, \theta_0) = \sum_{t=1}^N \sum_{i=1}^{|X|} P(x^t = x_i | y_t; \theta_0) \log P(x^t = x_i, y_t; \theta)$$

הוכחת אומדן ML למולטינום

מולטינום: $X = \{x_1, \dots, x_m\}$

אורך המדגם N
מספר הפעמים שנצפה x_i n_i

$$\sum_{i=1}^m n_i = N$$

הפרמטרים: $\theta = (p_1, \dots, p_m)$

$$\theta_{ML} = \operatorname{argmax}_{\theta} \prod_{i=1}^m p_i^{n_i}$$

נמקס את \log ההסתברות $\sum_{i=1}^m n_i \log p_i$. נמקסם:

$$\frac{1}{N} \sum_{i=1}^m n_i \log p_i = \sum_{i=1}^m \frac{n_i}{N} \log p_i \pm \sum_{i=1}^m \frac{n_i}{N} \log \frac{n_i}{N} = -D_{KL} \left(\frac{n_i}{N} \parallel p_i \right) + \underbrace{C}_{\text{constant}}$$

$$(D_{KL}(p \parallel q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \text{ - תזכורת})$$

קיבלנו שמקסימום יתקבל כאשר $\forall_i p_i = \frac{n_i}{N}$, וזה אומדן MLE למולטינום.

(*) נשים \heartsuit : מההוכחה קיבלנו, כאשר יש ביטוי מהצורה $\sum_{i=1}^m n_i \log p_i$, המקסימום מתקבל כאשר $\forall_i p_i = \frac{n_i}{N}$.

קעת נרצה למקסם את $Q(\theta, \theta_0)$ עבור מודל עירוב היסטוגרמות, כדי לקבל את משוואות צעד M - המשוואות שיקבעו את ערכי θ שממקסמים את $Q(\theta, \theta_0)$:

$$\begin{aligned} Q(\theta, \theta_0) &= \sum_{t=1}^N \sum_{i=1}^{|X|} \underbrace{P(x^t = x_i | t_y; \theta_0)}_{w_{ti} \text{ from the E-step}} \cdot \log P(x^t = x_i, y_t; \theta) = \\ &= \sum_{t=1}^N \sum_{i=1}^{|X|} w_{ti} \cdot \log \left[P(y_t | x^t = x_i; \theta) \cdot \underbrace{P(x^t = x_i; \theta)}_{P(x_i)} \right] = \\ &= \sum_{t=1}^N \sum_{i=1}^{|X|} w_{ti} \cdot \log P(x_i) + ??? \end{aligned}$$

ניזכר (V גודל המילון)

$$P(y_t | x_i) = \prod_{k=1}^V P(w_k | x_i)^{n_{tk}}$$

$$\log P(t_y|x_i) = \sum_{k=1}^V n_{tk} \log P(w_k|x_i)$$

ונציב בנוסחה:

$$\dots = \sum_{t=1}^N \underbrace{\sum_{i=1}^{|X|} w_{ti}}_{\text{constant}} \cdot \log P(x_i) + \sum_{i=1}^{|X|} \sum_{k=1}^V \underbrace{\sum_{t=1}^N w_{ti} \cdot n_{tk}}_{n_{ik}} \cdot \log P(w_k|x_i)$$

צריך למצוא ערכי $P(w_k|x_i)$ ו $P(x_i)$ שימקסמו את הביטוי לכל i ו k .
 נשים \heartsuit : כל אחד משני המחברים תלוי בסוג פרמטר אחר, ולכן ניתן למקסם כל אחד בנפרד.
 • נמקסם את המחובר השמאלי. לפי (*) נקבל מקסימום עבור

$$P(x_i) = \frac{\sum_{t=1}^N w_{ti}}{\underbrace{\sum_{j=1}^{|X|} \sum_{t=1}^N w_{tj}}_{=N}}$$

• נמקסם את המחובר הימני. נשים \heartsuit שבתוך כל איבר ב $\sum_{i=1}^{|X|}$ יש תלות רק בפרמטרים של ההתפלגות עבור x_i בנפרד. $P(w_k|x_i)$. כיוון שאין תלות בין ההתפלגויות המולטינומיות של x_i השונים, ניתן למקסם כל מחובר כזה בנפרד.

– כעת נמקסם מחובר עבור ערך i מסויים. רואים שכל מחובר הוא מהצורה $\underbrace{\sum_{k=1}^V n_{ik} \log P(w_k|x_i)}_{p_k}$.

לפי (*), ערך מקסימום מתקבל:

$$P(w_k|x_i) = \frac{n_{ik}}{\sum_{\ell=1}^V n_{i\ell}} = \frac{\sum_{t=1}^N w_{ti} n_{tk}}{\sum_{\ell=1}^V \sum_{t=1}^N w_{ti} \cdot n_{t\ell}} = \frac{\sum_{t=1}^N w_{ti} n_{tk}}{\sum_{t=1}^N w_{ti} \cdot n_t}$$