מקליד: עידן אריה
מרצה: פרופ' עידו דגן
תאריך: 2016-01-10

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

הסכימה הכללית של EM

המטרה: שערוך $\hat{\theta}$

אתחול: θ_0

איטרציה עד התכנסות:

צעד E: מחשבים $q\left(\theta_0\right) = p\left(x|y;\theta_0\right)$

צעד M: $\theta\left(q\right) = \underset{\theta}{\operatorname{argmax}} \sum_{x} p\left(x|y;\theta_0\right) \cdot \log p\left(x,y;\theta\right)$

$\theta_0 \leftarrow \theta\left(q\right)$

נפתח את משוואות E, M למודל עירוב היסטוגרמות

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

$$y = \left(y_1, \ldots y_N\right)$$

$$x = \left(x^1, \ldots x^N\right)$$

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

$$\log p\left(y;\theta\right) = \sum_{t=1}^{N} \log p\left(t_y;\theta\right)$$

ואז הסכימה הכללית למשוואות צעד הE והM תהיה:

צעד E: נחשב:

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

צעד M:

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

הוכחת אומדן ML למולטינום

מולטינום: $X=\{x_1,\dots x_m\}$

N	אורך המדגם
n_i	מספר הפעמים שנצפה x_i

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

הפרמטרים: $\theta=(p_1,\dots p_m)$

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

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

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

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

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

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

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

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

$$=\sum_{t=1}^{N}\sum_{i=1}^{|X|} w_{ti}\cdot\log\left[p\left(y_t\middle|x^t=x_i;\theta\right)\cdot\underbrace{P\left(x^t=x_i;\theta\right)}_{P(x_i)}\right]=$$

$$=\sum_{t=1}^{N}\sum_{i=1}^{|X|} w_{ti}\cdot\log p(x_i)+???$$

ניזכר(V גודל המילון)

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

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

ונציב בנוסחה:

$$\cdots=\sum_{t=1}^{N}\overbrace{\sum_{i=1}^{|X|}w_{t_i}}^{\text{constant}}\cdot\log p\left(x_i\right)+\sum_{i=1}^{|X|}\sum_{k=1}^{V}\underbrace{\sum_{t=1}^{N}w_{t_i}\cdot n_{tk}}_{n_{ik}}\cdot\log p\left(w_k|x_i\right)$$

צריך למצוא ערכי $p\left(w_k|x_i\right)$ ו-$p\left(x_i\right)$ שימקסמו את הביטוי לכל i ו-k.

נשים ♡: כל אחד משני המחוברים תלוי בסוג פרמטר אחר, ולכן ניתן למקסם כל אחד בנפרד.

- נמקסם את המחובר השמאלי. לפי (*) נקבל מקסימום עבור

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

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

  – כעת נמקסם מחובר עבור ערך i מסויים. רואים שכל מחובר הוא מהצורה $\sum_{k=1}^{V}n_{ik}\log\underbrace{p\left(w_k|x_i\right)}_{p_k}$. לפי (*), ערך מקסימום מתקבל:

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