

מטרה:

$$\theta_{ML} = \operatorname{argmax}_{\theta} \log P(y; \theta) = \operatorname{argmax}_{\theta} \log \sum_x P(x, y; \theta)$$

סכימת אופטימיזציה איטרטיבית למציאת מקסימום לוקאלי

Alternate Maximization

נסתכל על פונקציה $g(\psi)$, כאשר ψ קבוצת הפרמטרים. רוצים למקסם - אבל אין פתרון אנליטי, או אלגוריתם למציאת מקסימום.

כלומר: קשה/לא ניתן למצוא $\operatorname{argmax}_{\psi} g(\psi)$

אבל **נניח** שניתן לחלק את ψ לשתי קבוצות θ ו ϕ - כך שנרשום $g(\theta, \phi)$ ומתקיים שכאשר מקבעים קבוצת פרמטרים אחת אז ניתן/קל למצוא מקסימום g עבור קבוצת הפרמטרים השנייה.

כלומר: • בקיבוע θ_0 ניתן למצוא $\operatorname{argmax}_{\phi} g(\phi, \theta_0)$

• בקיבוע ϕ_0 ניתן למצוא $\operatorname{argmax}_{\theta} g(\phi_0, \theta)$

g כזו ניתן להגדיר אלגוריתם איטרטיבי למציאת מקסימום לוקאלי:

אלגוריתם AltMax

אתחול: נקבע θ_0

1. עבור θ_0 מקובע נמצא $\phi_0 = \operatorname{argmax}_{\phi} g(\phi, \theta_0)$

2. עבור ϕ_0 מקובע נמצא $\theta^* = \operatorname{argmax}_{\theta} g(\phi_0, \theta)$

3. $\theta_0 \leftarrow \theta^*$ ונחזור ל1

מובטח שבכל צעד בכל איטרציה, ערך g לא ירד, עד התכנסות למקסימום לוקאלי.

ואם אין חלוקה?

נרצה להפעיל את סכימת AltMax לפונקציה f שלא ניתן לחלק דדדדאת הפרמטרים שלה כנ"ל.

כלומר: נחפש $\operatorname{argmax}_{\theta} f(\theta)$ כאשר אין פתרון ישיר למקסום.

נחפש פונקציה $g(\theta, \phi)$, כאשר ϕ זו קבוצה של פרמטרי עזר, כך ש $f(\theta) = \max_{\phi} g(\theta, \phi)$. ולכן יתקיים:

$$\operatorname{argmax}_{\theta} f(\theta) = \operatorname{argmax}_{\theta} \max_{\phi} g(\theta, \phi)$$

כעת ניתן להפעיל על g את אלגוריתם AltMax שכתבנו.

נשים \heartsuit : בצעד 1 ערך $g(\phi_0, \theta_0)$ מתלכד עם ערך $f(\theta_0)$, ולכן גם ערך $f(\theta_0)$ עולה בכל איטרציה, עד להתכנסות למקסימום לוקאלי של $f(\theta)$

נרצה להפעיל את סכימת AltMax על פונקציית הנראות: נחפש פונקציה עם פרמטרי עזר, שחסומה מלמעלה ע"י פונקציית הנראות:

$$\log P(y; \theta) = \log \sum_x P(x, y; \theta) = \log \sum_x q(x) \cdot \frac{P(x, y; \theta)}{q(x)}$$

כאשר $q(x)$ היא התפלגות מעל X שערכיה מהווים את אוסף משתני העזר ϕ (AltMax). נפעיל את אי-שוויון ג'נסן, בהסתכלות על $\frac{P(x, y; \theta)}{q(x)}$ כפונקציה מעל x , והסכום הוא תוחלת של הפונקציה לפי התפלגות $q(x)$ (כזכור - \log קעורה):

$$\dots \geq \sum_x q(x) \log \frac{P(x, y; \theta)}{q(x)} \triangleq F(\theta, q)$$

הפונקציה הזו F נקראת "פונקציית האנרגיה החופשית".

קיבלנו: עבור כל התפלגות שהיא q (בהנחה ש $\forall_x q(x) > 0$) מתקיים:

$$\log P(y; \theta) \geq F(\theta, q)$$

נוכל להשתמש בסכימת AltMax אם תמיד קיים q שעבורו מתקיים שוויון. כלומר שיתקיים:

$$\log P(y; \theta) = \max_q F(\theta, q)$$

כעת נוכיח שמתקיים שוויון עבור q מסויים, ונראה מתי הוא מתקיים: נסתכל על ההפרש בין הביטויים ונבדוק אם ומתי הוא מתאפס:

$$\log P(y; \theta) - F(\theta, q) \stackrel{?}{=} 0$$

נגרום לזה להראות כמו D_{KL} (תזכורת - $D_{KL}(p||q) = \sum_x p(x) \cdot \log \frac{p(x)}{q(x)}$) - על ידי כך שנכפיל את שני הצדדים ב $\sum_x q(x)$:

$$\sum_x q(x) \cdot \log p(y; \theta) - \sum_x q(x) \log \frac{P(x, y; \theta)}{q(x)} = \sum_x q(x) \log \frac{q(x)}{\frac{P(x, y; \theta)}{P(y; \theta)}} = D_{KL}[g(x) || P(x|y; \theta)]$$

כזכור, $D_{KL} = 0$ כאשר שתי ההתפלגויות שוות, כלומר $q(x) = P(x|y; \theta)$. כעת ניתן לייצג את θ_{ML} באופן הבא:

$$\theta_{ML} = \operatorname{argmax}_{\theta} \log P(y; \theta) = \operatorname{argmax}_{\theta}$$

כעת קיבלנו ייצוג של הנראות:

$$\log P(y; \theta) = \max_q F(\theta, q) = F(\theta, P(x|y; \theta))$$

כלומר: צריך למצוא זוג של θ, q שממקסם את, ובפרט ערך θ בזוג הוא θ_{ML} עבור הנראות.

שלבי אלגוריתם EM

הפעלת סכימת AltMax על פונקציית הנראות, כאשר $F(\theta, q(x))$ היא פונקציית העזר:

אתחול: נאתחל את θ_0

צעד 1 - E-step

חישוב משוואות צעד-E:

$$q(\theta_0) = \operatorname{argmax}_q F(\theta, q) = P(x|y; \theta_0)$$

זו פונקציית הסיווג - נחשב אותה לכל ערך $x \in X$, עבור y שנתון בתצפית ו θ_0 שקיבענו (מקביל ל w_{ti} שראינו בעירוב היסטוגרמות)

צעד 2 - M-step

בהינתן $q(\theta_0)$, נמצא θ שממקסם את $F(\theta, q(\theta_0))$:

$$\begin{aligned} \theta \left(\underbrace{q}_{\substack{\text{fixed} \\ \text{to } q(\theta_0)}} \right) &= \operatorname{argmax}_\theta F(\theta, q) = \operatorname{argmax}_\theta \left[\sum_x q(x) \log P(x, y; \theta) - \overbrace{\sum_x q(x) \log q(x)}^{\text{fixed (for a fixed } q)}} \right] = \\ &= \operatorname{argmax}_\theta \left[\sum_x P(x|y; \theta_0) \cdot \log P(x, y; \theta) \right] \triangleq \operatorname{argmax}_\theta Q(\theta, \theta_0) \end{aligned}$$

כאשר Q היא פונקציית עזר Auxiliary Function.

בשלב זה: צריך למצוא מיקסום לפי θ של Q , בהתאם למשוואות הסיווג והנראות של המודל שאיתו עובדים.

פעמים רבות - יש פתרון אנליטי.

(סכום של \log ים קל יותר למקסם)