מקליד: עידן אריה
מרצה: פרופ' עידו דגן
תאריך: 2016-01-03

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

מטרה:

$$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \log p\left(y;\theta\right) = \underset{\theta}{\operatorname{argmax}} \log \sum_{x} p\left(x,y;\theta\right)$$

סכימת אופטימיזציה איטרטיבית למציאת מקסימום לוקאלי

Alternate Maximization

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

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

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

כלומר:
- בקיבוע θ_0 ניתן למצוא $\underset{\phi}{\operatorname{argmax}} g\left(\phi,\theta_0\right)$
- בקיבוע ϕ_0 ניתן למצוא $\underset{\theta}{\operatorname{argmax}} g\left(\phi_0,\theta\right)$

לg כזו ניתן להגדיר אלגוריתם איטרטיבי למציאת מקסימום לוקאלי:

אלגוריתם AltMax

אתחול: נקבע θ_0

1. עבור θ_0 מקובע נמצא $\phi_0 = \underset{\phi}{\operatorname{argmax}} g\left(\phi,\theta_0\right)$
2. עבור ϕ_0 מקובע נמצא $\theta^* = \underset{\theta}{\operatorname{argmax}} g\left(\phi_0,\theta\right)$
3. $\theta^* \leftarrow \theta_0$ 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 $\underset{\theta}{\operatorname{argmax}} f\left(\theta\right)$ כאשר אין פתרון ישיר למקסום.

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

$$\underset{\theta}{\operatorname{argmax}} f\left(\theta\right) = \underset{\theta}{\operatorname{argmax}} \max_{\phi} g\left(\theta,\phi\right)$$

כעת ניתן להפעיל על g את אלגוריתם AltMax שכתבנו.

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

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 מAltMax).
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):

$$\cdots\geq\sum_x q(x)\log\frac{p(x,y;\theta)}{q(x)}\triangleq F(\theta,g)$$

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 - $D_{KL}(p\|q)=\sum\limits_x p(x)\cdot\log\frac{p(x)}{q(x)}$) - על ידי כך שנכפיל את שני הצדדים ב$\sum\limits_x q(x)$:

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

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

$$\theta_{ML}=\underset{\theta}{\operatorname{argmax}}\log p(y;\theta)=\underset{\theta}{\operatorname{argmax}}$$

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 $F\left(\theta,q\left(x\right)\right)$ היא פונקציית העזר:

אתחול: נאתחל את θ_0

צעד 1 - E-step

חישוב משוואות צעד-E:

$$q\left(\theta_0\right)=\underset{q}{\text{argmax}}F\left(\theta,q\right)=p\left(x|y;\theta_0\right)$$

זו פונקציית הסיווג - נחשב אותה לכל ערך $x\in X$, עבור הy שנתון בתצפית וθ_0 שקיבענו(מקביל לw_{ti} שראינו בעירוב היסטוגרמות)

צעד 2 - M-step

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

$$\theta\left(\overbrace{q}^{\substack{\text{fixed}\\ \text{to } q\left(\theta_0\right)}}\right)=\underset{\theta}{\text{argmax}}F\left(\theta,q\right)=\underset{\theta}{\text{argmax}}\left[\sum_x q\left(x\right)\log p\left(x,y;\theta\right)-\overbrace{\sum_x q\left(x\right)\log q\left(x\right)}^{\text{fixed(fox a fixed } q)}\right]=$$

$$=\underset{\theta}{\text{argmax}}\left[\sum_x P\left(x|y;\theta_0\right)\cdot\log p\left(x,y;\theta\right)\right]\triangleq\underset{\theta}{\text{argmax}}Q\left(\theta,\theta_0\right)$$

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 logים קל יותר למקסם)