מקליד: עידן אריה
מרצה: פרופ' עידו דגן
תאריך: 2015-11-01

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

התפלגות משותפת של מ"מ X,Y - $p\left(X,Y\right)$

התפלגות שולית

$$p_X\left(x\right)=\sum_{y\in Y}p\left(x,y\right)$$

$$p_Y\left(y\right)=\sum_{x\in X}p\left(x,y\right)$$

טריוויאלית:

$$\sum_{x}p_X\left(x\right)=1$$

תוחלת של פונקציה מעל התפלגות משותפת

$$E\left[g\left(x,y\right)\right]=\sum_{x,y}g\left(x,y\right)\cdot p\left(x,y\right)$$

לדוגמה

תוחלת של סכום ערכי משתנים מקריים בהתפלגות משותפת שווה לסכום התוחלות של שני המשתנים(ע"פ ההתפלגויות השוליות).

הוכחה

$$E\left[X+Y\right]=\sum_{x}\sum_{y}\left(x+y\right)\cdot p\left(x,y\right)=$$

$$=\sum_{x}\sum_{y}x\cdot p\left(x,y\right)+\sum_{x}\sum_{y}y\cdot p\left(x,y\right)=$$

$$=\sum_{x}x\cdot p\left(x\right)+\sum_{y}y\cdot p\left(y\right)=E\left[X\right]+E\left[Y\right]$$

משתנה מקרי בלתי תלוי

מ"מ X וY נקראים בלתי תלויים(ב"ת) אם לכל $x\in X$ ו$y\in Y$ מתקיים:

$$p\left(x,y\right)=p_X\left(x\right)p_Y\left(y\right)$$

התפלגות מותנית של מ"מ $p_{X|Y}(x|y)$

פונקציית הסתברות שמגדירה

$$p_{X|Y}(x|y) = P(\{X=x\}|\{Y=y\}) = \frac{P(\{X=x, Y=y\})}{P(\{Y=y\})} = \frac{p(x,y)}{p(y)}$$

מתקיים לכל y מסויים:

$$\sum_x p(x|y) = 1$$

אמדן (Estimation)

הקדמה - שימוש במודלים הסתברותיים

נגדיר מודל הסתברותי לתופעה שבה מתעניינים, למטרות שונות:

- הסבר - פענוח, סיווג. לדוגמה - להחליט באיזה נושא הכתבה לפי המילים שיש בה
- חיזוי - מזג אוויר, בורסה
- יצירה - יצירת דיבור, תרגום

נגדיר אוסף מ"מ ומהם הפרמטרים שלפיהם נקבעים ערכי המשתנים.

לדוגמה: מטבע - כגון הסתברות המילה "ילד" בטקסט - נמדל כבינום.

בד"כ נסתכל על ההתנהגות "הגנרטיבית"של המודל שמייצרת את את התצפיות בניסוי

לדוגמה: מודל הסתברות מילה "ילד" כבינום.

תהליך גנרטיבי אפשרי: מעבר על כל מקום בטקסט, ועבורו מגרילים בהסתברות p האם תופיע המילה "ילד" או מילה אחרת.

נשים ♡: זהו מודל מאוד פשטני, ויש מודלים שממדלים יותר טוב - למשל מסתכלים על ההקשר של המשפט. אבל בד"כ נרצה להגביל את כמות המשתנים כדי לפשט את המודל, ולכן המודלים שלנו יהיו מקורבים.

סימון מקובל

θ אוסף הערכים להסתברויות שמוגדרות במודל - הפרמטרים של המודל

לדוגמה - במשתנה בינומי $\theta = \{p\}$

שערוך מודל

תהליך אומדן ערכי ההסתברויות שבו θ. יתבצע ע"י דגימה סטטיסטית.

למשל: אומדן p למטבע: נתיל 100 הטלות ונניח שקיבלנו 70 הצלחות $\Longleftarrow$ נשערך $p = 0.7$.
למה?

אומדן נראות מקסימלית(Maximum Likelihood Estimation)

נראות(Likelihood)

אינטואיציה: בהנתן שאני יודע מה ערכי הפרמטרים, מהי ההסתברות לתצפית מסויימת?

עבור תצפית מסויימת, הנראות של התצפית היא ההסתברות לקבל אותה על פי ערכי פרמטרים נתונים של המודל.

נסמן: $p\left(\overbrace{O}^{\text{observation}};\theta\right)$

לדוגמה: בבינום: $p\left(70,100;p\right)$(70 הצלחות מתוך 100 נסיונות עבור הפרמטר p)
הפונקציה הזו היא פונקציה של p - לכל p אפשרי היא קובעת את ההסתברות לקבלת התצפית עבור ערך זה של p

הרציונאל אומדן נראות מקסימלית

מכיוון שאנחנו יודעים את התצפית ורוצים לחשב לפיה את הפרמטרים, נרצה למצוא θ כך ש$p\left(O,\theta\right)$ מקסימלי

אומדן נראות מקסימלי

נבחר כאומדן נראות מקסימלי(MLE):

$$p_{MLE}\left(\langle 70,100\rangle\right)=\text{argmax}_p p\left(70,100;p\right)=0.7$$

באופן כללי

$$\theta_{ML}\left(\overbrace{O}^{\text{given observation}}\right)=\text{argmax}_\theta p\left(\overbrace{O}^{\text{constant}};\overbrace{\theta}^{\text{parameter}}\right)$$

נוכיח - לבינום $p_{MLE}\left(\left\langle\overbrace{n}^{\textbf{attempts}},\overbrace{k}^{\textbf{successes}}\right\rangle\right)=\frac{k}{n}$

עבור k,n נתונים בתצפית(קבועים):
פונקציית הנראות: $p\left(k,n;p\right)=\binom{n}{k}p^k\left(1-p\right)^{n-k}$
אנו רוצים למצוא את ערך p שימקסם את פונקציית הנראות: לפונקציית הבינום ניתן למצוא פיתרון אנליטי ע"י גזירה.
יהיה נוח יותר לגזור את הlog של פונקציית הנראות - ובגלל מונוטוניות log, p שימקסם את הlog ממקסמם את הפונקציה המקורית. נסמן $L-\log\left(\text{likelihood}\right)$

$$L\left(p\right)=\log p\left(k,n;p\right)=\log\left(\binom{n}{k}p^k\left(1-p\right)^{n-k}\right)=\log\binom{n}{k}+k\log p+\left(n-k\right)\log\left(1-p\right)$$

נגזור(תזכורת $\left(\log x\right)'=\frac{1}{x}$) ונשווה לאפס:

$$\frac{\mathrm{d}L\left(p\right)}{\mathrm{d}p}=\frac{k}{p}-\frac{n-k}{1-p}=0$$

$$k-\cancel{kp}=np-\cancel{kp}\implies p=\frac{k}{n}$$

הערה: כדי להוכיח שזו נקודת מקסימום, צריך להראות שהפונקציה קעורה($\cap$[1]): נגזרת שנייה שלילית בכל התחום של p.

ההצדקה לMLE

באיזה מובן MLE הוא הנכון?

אינטואיציה: נניח שיש לנו מטבע שהפרמטר האמיתי שלו הוא 0.7, ואנו רוצים למצוא את זה. אם נטיל אותו $n=100$ פעמים ונקבל נגיד $k=68$. במקרה כזה $p_{MLE}=0.68$. אם נגדיל ל$n=100000$ נקבל אולי $p_{MLE}=0.705$. אפשר להוכיח שככל שנגדיל את n, p_{MLE} יתכנס לערך האמיתי p - זהו אומדן חסר הטיה[2]

ניתן להוכיח: עבור תצפית שנוצרות על ידיד הפרמטרים של המודל(האמייתיים - אלו שאותם מנסים לשערך) מתקיים שאומדן MLE מתכנס לערכי הפרמטרים עם הגדלת המדגם(שואף ל∞).

זה נקרא: אומד עקבי אסימפטוטית Asymptotically consistent

[1] כאשר מסתכלים מלמטה $\cap$ זה קעור ו$\cup$ זה קמור
[2] אומדן חסר הטיה - אומדן שככל שמגדילים את גודל המדגם מתכנס לערך האמיתי