פילוסופיה בייזיאנית

<u>תזכורת</u>: בשבוע שעבר דיברנו על חוק ב"ס-

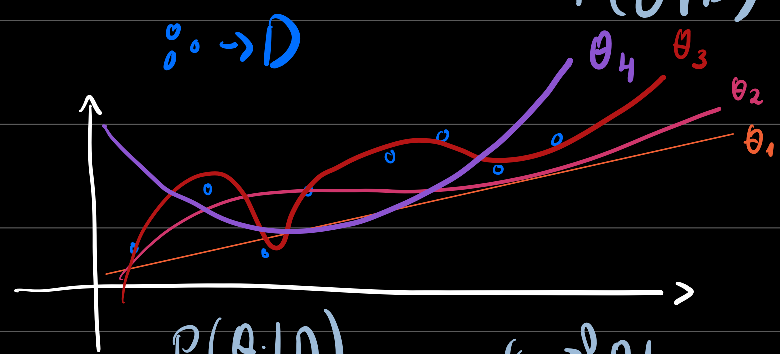
$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)\cdot P(B|A)}{P(B)}$$

"Law of Inverse probability"

<u>הפילוסופיה הבייזיאנית:</u>

בהינתן נתונים D ופרמטר Θ שרוצים לחזות,
נחשב את $P(\Theta|D)$

לדוגמא-



נחשב את $P(\theta_i|D)$

לכל $i=1,2,3,\ldots$

דוגמא: נתונים $D=(1,0,0,1,0,0,0,1)$ תוצאות
של הטלת מטבע.

θ - הסיכוי שהמטבע ייצא עץ.

$1-\theta$ " " פלי.

נוכל בקלות לחשב את $P(\theta|D)$ לכל $\theta\in[0,1]$.

הפילוסופיה הקלאסית:

בהינתן נתונים D ופרמטר θ נחשב את θ^* שהוא הפרמטר המתאים ביותר ל-D תחת פונקציית הפסד מוגדרת (loss)

$$\theta^* = \operatorname{argmin}_{\theta} L(\theta, D)$$

יתרונות הגישה הבייסיאנית:

1. אין דרגת חופש לגבי $L(\theta,D)$ (או כל דבר אחר).
2. אופטימלי - "יותר טוב מכל אלגוריתם אחר".

<u>דוגמא</u>- θ הסיכוי ל-0 $D=(\underbrace{1,0,1,0,0,\dots}_{N})$

$1-\theta$ הסיכוי ל-1 N פריטים ב"ת.

אם היינו יודעים את θ, היה קל לחשב

נוטציה שמבטאת פונקציה ופרמטר. זו פונקציה של D עם פרמטר θ. <u>לא</u> הסתברות מותנית

$$P(D;\theta)=\theta^{\#0}\cdot(1-\theta)^{\#1}$$

בגישה הקלאסית, משתמשים הרבה בשיטת הנראות המקסימלית - Maximum Likelihood Estimation

$$\theta^*=\arg\max_\theta P(D;\theta)$$

במקרה שלנו-

$$\theta^*=\arg\max_\theta \theta^{\#0}\cdot(1-\theta)^{\#1}$$

$$=\arg\max_\theta \underbrace{\#0\cdot\log(\theta)+\#1\cdot\log(1-\theta)}_{=J(\theta)}$$

עכשיו קל לגזור את J ולהשוות ל-0

$$\frac{\partial J}{\partial \theta} = \frac{\#0}{\theta} - \frac{\#1}{1-\theta} = 0$$

$$\Downarrow$$

$$\frac{\#0}{\theta} = \frac{\#1}{1-\theta}$$

$$\Downarrow$$

$$\frac{\theta}{\#0} = \frac{1-\theta}{\#1}$$

$$\Downarrow$$

$$\frac{\theta}{\#0} = \frac{1-\theta}{N-\#0}$$

$$\Downarrow$$

$$\theta = \frac{\#0}{N}$$

בגישה הבייזיאנית - $P(\theta|D) = \frac{P(\theta)\cdot P(\theta,D)}{P(D)}$

- לפי גישה זו, θ הוא משתנה מקרי. מותר לכתוב $P(\theta|D), P(\theta)$ וכו'.
- בגישה הקלאסית θ הוא פרמטר, לכן <u>אסור</u> לכתוב $P(\theta|D), P(\theta)$ משום שלא מוגדר מרחב הסתברות.

בגישה השכיחותית

$P(X)$ זו השכיחות הצפויה למאורע X אם נחזור על הניסוי אינסוף פעמים.

- אם נטיל מטבע הוגן אינסוף פעמים, נצפה לקבל שחצי מההטלות יצאו עץ וחצי פלי.
- מה ההסתברות שהמטבע הוגן? השאלה הנ"ל לא מוגדרת היטב בגישה השכיחותית.

בגישה הבייזיאנית

$P(X)$ מבטא את דרגת האמונה שלי שיקרה המאורע X.

- בגישה הבייזיאנית השאלה "מה הסיכוי שהמטבע הוגן?" היא שאלה קבילה.

בגישה זו, חישוב ההסתברות ל-D "נעשה כך-

$$P(D|\theta) = \prod_{i=1}^{N} P(x_i|\theta) = \theta^{\#0} \cdot (1-\theta)^{\#1}$$

במקרה זה, התוצאות בלתי תלויות ב~~בניסוי~~ θ

נניח כי $\theta \sim U[0,1]$, אזי

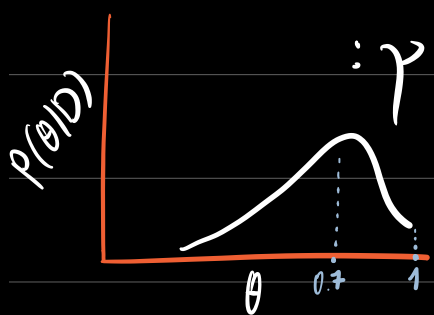
$$P(\theta|D) = \frac{\overset{1}{\overset{\shortparallel}{P(\theta)}}\, P(\theta|D)}{P(D)} = \frac{1 \cdot \theta^{\#0}(1-\theta)^{\#1}}{\int_{\theta'} P(\theta') \cdot P(D|\theta')}$$

נוסחאת ההסתברות השלמה

$P(A) = \sum_B P(A,B) = \sum_B P(B) \cdot P(A|B)$

$$= \frac{1 \cdot \theta^{\#0}(1-\theta)^{\#1}}{\int_{\theta'} 1 \cdot \left(\theta'^{\#0} \cdot (1-\theta')^{\#1}\right) d\theta'}$$

אם $N=1000$, $\#0=700$, $\#1=300$, נקבל פונק' צפיפות של θ שנראית בערך כך:



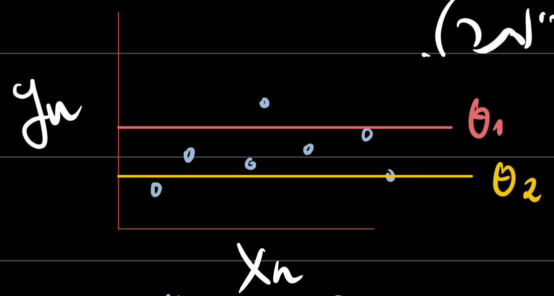
חסרונות הגישה הבייסיאנית:

1. מאיפה מגיע בכלל $P(\theta)$?
2. החישוב של $P(\theta|D)$ עשוי להיות קשה. בדוגמה הקודמת צריך לדעת לחשב את $\int_{\theta'} \left(\theta'^{\#0} \cdot (1-\theta')^{\#1}\right) d\theta'$

לא סיפור פשוט, רק מי שלמד אינפי 3 יודע...

אופטימליות הגישה הבייזיאנית:

יש לנו אוסף מדידות של מס' מאומתים לקורונה בכל יום. אנחנו רוצים למצוא את תוחלת מס' החולים ביום (כלומר הקו האופקי שמסביר את הנתונים בצורה הטובה ביותר).



אוניברסיטה א' משתמשת באלג' א' ומקבלת θ_1^*

" ב' " " ב' " θ_2^*

איך נמדוד איזו אוניברסיטה טובה יותר?

רעיון א' - נחביא חלק מהדאטה ונבחן מי יותר טוב על הדאטה שהאוניברסיטאות לא ראו.

רעיון ב' - נייצר נתונים מפונקציה ידועה θ^{GT} ונמדוד את האוניברסיטאות לפי הקירבה ל- θ^{GT}.

הגדרה קלאסית של טיב שיערוך -

$$\theta^{GT} \Rightarrow \{X_i\} \quad P(X_i ; \theta^{GT}) \Rightarrow \hat{\theta}(\{X_i\})$$

מייצרים דאטא $\{X_i\}$ בעזרת פרמטר θ^{GT}

מקבלים את הדוגמאות $\{X_i\}$ ומפעילים את האלג' $\hat{\theta}$ לקבלת שיערוך

הגדרה -

$$MSE(\hat{\theta}) = \mathbb{E}_{X \sim P(X; \theta^{GT})} |\theta^{GT} - \hat{\theta}(X)|^2$$

$$BMSE(\hat{\theta}) = \mathbb{E}_{\theta^{GT}, X} |\theta^{GT} - \hat{\theta}(X)|$$

טענה: $\hat{\theta}(D) = \mathbb{E}_{\theta}(\theta | D)$ הוא המשערך הטוב ביותר האפשרי לפי שיערוך BMSE.