האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

אלגוריתמים בביו' חישובית

76558

למידת פרמטרים במודלים מרקוביים חבויים

תומי קפלן

6/2/2024

למידת פרמטרים במודל מרקובי חבוי

דיברנו על חישובי ניראות והסקה. אבל מאיפה המודל?

- לימוד **מבנה** – נגענו טיפה. מורכב. יצירתיות וסבלנות.

- לימוד **פרמטרים** – נניח סט אימון של רצפים ב"ת

$$LL(\theta : D) = \log P(\overrightarrow{X^1}, \ldots, \overrightarrow{X^N} \mid \theta) = \sum_{j=1}^{N} \log P(\overrightarrow{X^j} \mid \theta)$$

θ סט הפרמטרים של המודל: transitions : $\tau_{k,l}$

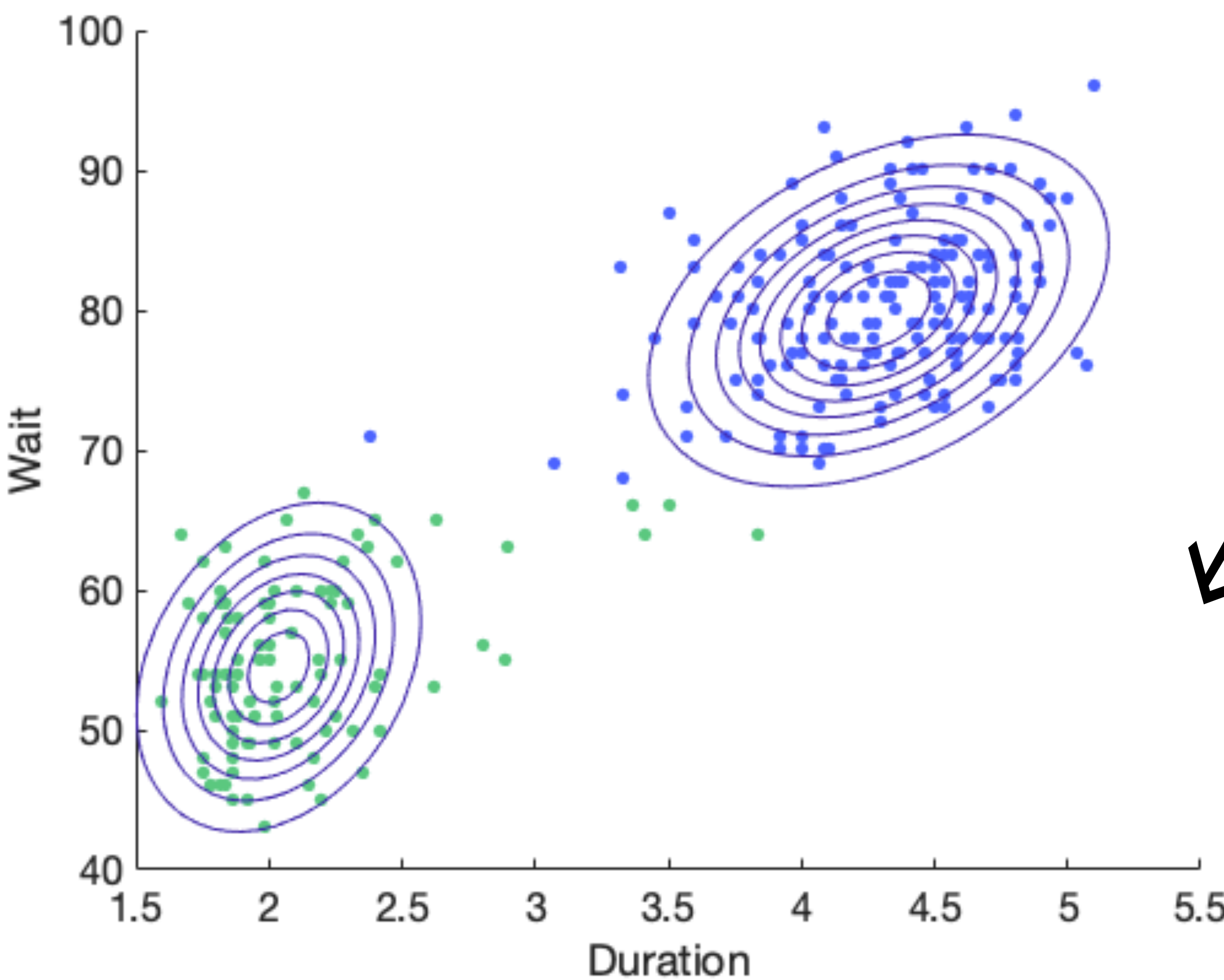
emissions : $e_L(k)$

- אז איך לומדים את הפרמטרים האופטימליים?
  [הרי המצבים החבויים חבויים]

לא מאוד שונה מבעיות שראינו בעבר

- למשל, בעיית הקלאסטרינג – קיבוץ דוגמאות לסוגים

Gaussian Mixture Model



- תיוג הדוגמאות חסר
- אלגוריתם K-means:

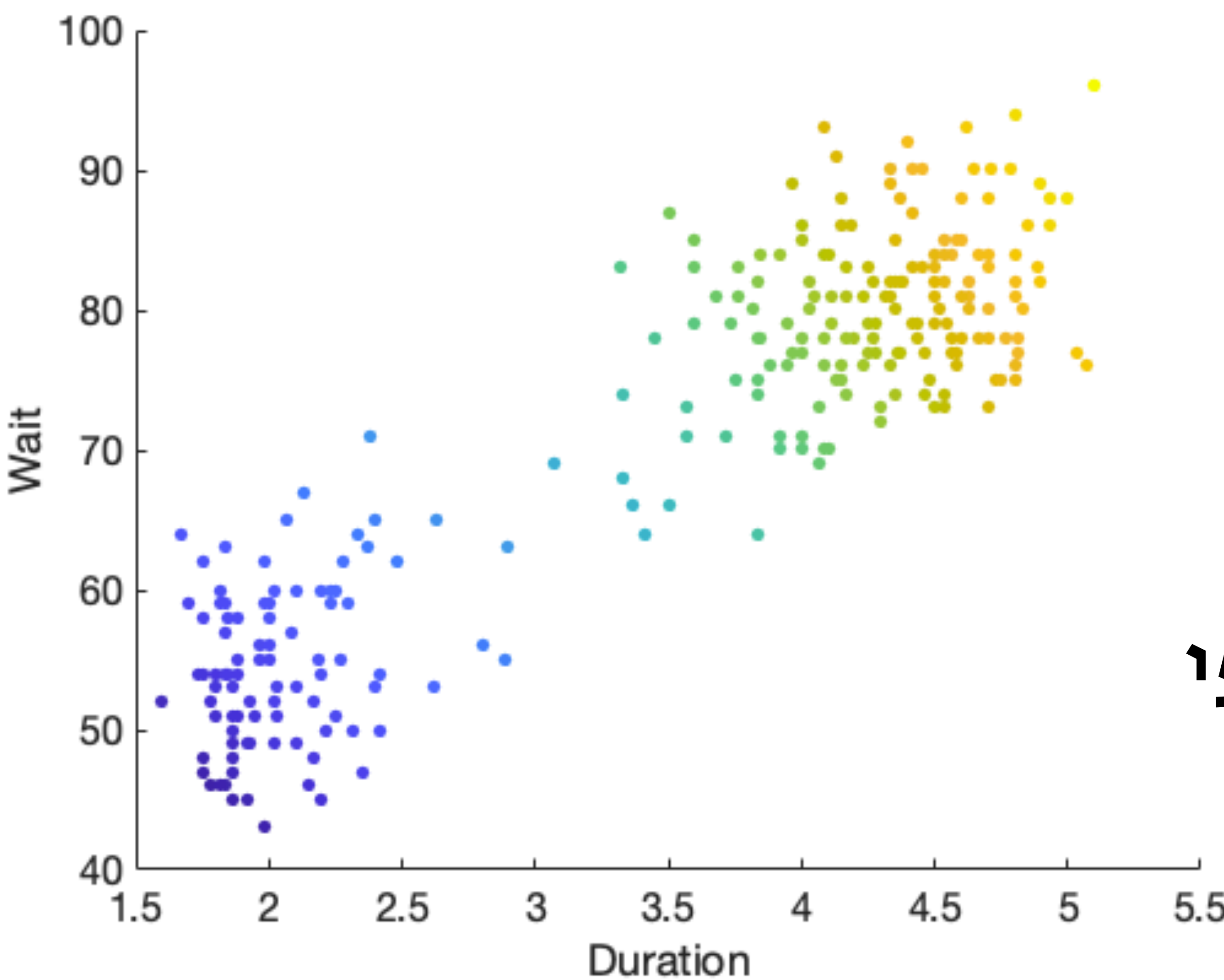
שיוך דוגמא לקלאסטר

שערוך המרכזים מחדש

השלמת המידע ע״ס התצפיות והפרמטרים הקיימים

קלאסטרינג בגירסא רכה

- כל נקודה משוייכת לכל מרכז, בהסתברות כלשהי [ההסת' א-פוסטריורי, להגיע מכל אחד מהמרכזים]



- אלגוריתם רך:

הניראות והפוסטריור לפי כל קלאסטר

ממוצע ושונות של כל קבוצה, ממושקלים לפי השיוך

נחפש אלגוריתם דומה למודלים מרקוביים חבויים

למידת פרמטרים במודל מרקובי חבוי

- נניח לרגע שהדאטה היה מתוייג כולו

$$LL(\theta : D^*) = \sum_{j=1}^{N} \log P(\overrightarrow{X^j}, \overrightarrow{S^j} \mid \theta)$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[\log P(S_i^j \mid S_{i-1}^j) + \log P(X_i^j \mid S_i^j) \right]$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[\log \tau_{S_{i-1}^j, S_i^j} + \log \mathsf{e}_{S_i^j}(X_i^j) \right]$$

$$= \sum_{k,l} N_{k,l} \cdot \log \tau_{k,l} + \sum_{k,x} N_{k,x} \cdot \log \mathsf{e}_k(x)$$

- קל לגזור ולמצוא את θ שממקסמת את לוג הנראות

למידת פרמטרים במודל מרקובי חבוי

- אבל אין לנו מידע מתויג אודות המסלולים $\{\overrightarrow{S^j}\}$
- הניראות דורשת סכימה על כל ההשמות האפשריות

$$L(\theta : D^*) = \prod_{j=1}^{N} P(\overrightarrow{X^j} \,|\, \theta) = \prod_{j=1}^{N} \sum_{\overrightarrow{S^j}} \prod_{i=1}^{N_j} \left[\tau_{S^j_{i-1}, S^j_i} \cdot \mathsf{e}_{S^j_i}(X^j_i) \right]$$

- ואנו מחפשים את θ הממקסמת את [לוג] הניראות

$$\theta^* = \arg\max_{\theta} \; L(\theta : D) = \arg\max_{\theta} \; LL(\theta : D) = \sum_{j=1}^{N} \log P(\overrightarrow{X^j} \,|\, \theta)$$

מה נעשה?

- **מציאת מסלולי ויטרבי $\Leftrightarrow$ שערוך פרמטרים**

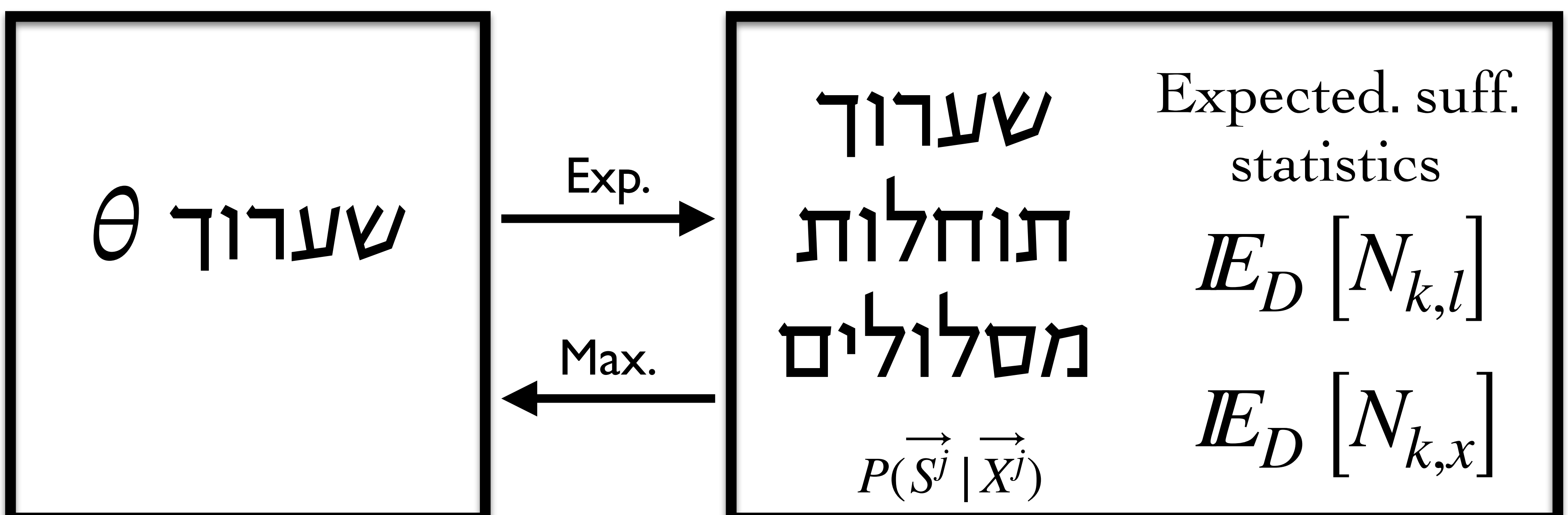
- **אופטימיזציה מבוססת גרדיאנט**
נאתחל את θ ונבצע שינויים קטנים לשיפור הניראות,
עד להתכנסות [לאופטימום לוקאלי]

- **אלגוריתם באום-וולש** [Baum-Welch]
אלגוריתם דו-צדדי.
צעד E משערך תוחלות על המסלולים החבויים*
צעד M משערך את הפרמטרים על פיהם

*ועל הסטטיסטי המספיקים

Expectation-Maximization

- **אלגוריתם באום-וולש** [Baum-Welch]
אלגוריתם דו-צדדי.
צעד E משערך תוחלות על המסלולים החבויים*
צעד M משערך את הפרמטרים על פיהם



שערוך MLE לפרמטרים

- ניזכר בסטטיסטי המספיקים לאומדי MLE

$$\tilde{\tau}_{k,l} = \frac{N_{k,l}}{\sum_m N_{k,m}} \qquad \tilde{e}_k(x) = \frac{N_{k,x}}{\sum_y N_{k,y}}$$

- כאשר

$$N_{k,l} = \sum_{j=1}^{N} \sum_{i=2}^{n_j} \mathbb{1}\left\{S_{i-1}^j = k, S_i^j = l\right\} \qquad N_{k,x} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} \mathbb{1}\left\{S_i^j = k, X_i^j = x\right\}$$

- כאשר $\{S^j\}$ לא ידוע, נסתפק בתוחלות

$$\tilde{N}_{k,l} = \sum_{j=1}^{N} \sum_{i=2}^{n_j} E_\theta \mathbb{1}\left\{S_{i-1}^j = k, S_i^j = l\right\} \qquad \tilde{N}_{k,x} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} E_\theta \mathbb{1}\left\{S_i^j = k, X_i^j = x\right\}$$

שערוך MLE לפרמטרים

- כאשר $\{S^j\}$ לא ידוע, נסתפק בתוחלות

$$\tilde{N}_{k,l} = \sum_{j=1}^{N} \sum_{i=2}^{n_j} E_\theta \mathbb{1}\left\{S_{i-1}^j = k, S_i^j = l\right\}$$
$$= \sum_{j=1}^{N} \sum_{i=2}^{n_j} P(S_{i-1}^j = k, S_i^j = l \mid \overrightarrow{X}^j)$$
$$= \sum_{j=1}^{N} \sum_{i=2}^{n_j} \frac{\left[F_k(i-1) \cdot \tau_{k,l} \cdot \mathsf{e}_l(X_i) \cdot B_l(i)\right]}{P(\overrightarrow{X}^j)}$$

$$\tilde{N}_{k,x} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} E_\theta \mathbb{1}\left\{S_i^j = k, X_i^j = x\right\}$$
$$= \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[P(S_i^j = k \mid \overrightarrow{X}^j) \cdot P(X_i^j = x \mid S_i^j = k, \overrightarrow{X}^j)\right] = \sum_{j=1}^{N} \sum_{i:X_i^j = x} P(S_i^j = k \mid \overrightarrow{X}^j)$$
$$= \sum_{j=1}^{N} \sum_{i:X_i^j = x} \frac{F_k(i) \cdot B_k(i)}{P(\overrightarrow{X}^j)}$$

אלגוריתם באום-וולש

- אתחול θ (רנדומית)
- בכל איטרציה t
  - לכל רצף X^j חשבו את הפורוורד והבקוורד (בעזרת θ הנוכחית)

  - שלב E:

$$\tilde{N}_{k,x} = \sum_{j=1}^{N} \sum_{i:X_i^j=x} \frac{F_k(i)\cdot B_k(i)}{P(\vec{X}^j)} \qquad \tilde{N}_{k,l} = \sum_{j=1}^{N}\sum_{i=2}^{n_j}\left[F_k(i-1)\cdot\tau_{k,l}\cdot e_l(X_i)\cdot B_l(i)\right]$$

  - שלב M:

$$\tilde{e}_k(x) = \frac{N_{k,x}}{\sum_y N_{k,y}} \qquad \tilde{\tau}_{k,l} = \frac{N_{k,l}}{\sum_m N_{k,m}}$$

  - עיצרו אם השיפור בלוג הנראות (או אם השינוי ב- θ) קטן מספיק

הבסיס התיאורטי ל-EM

- נסתכל על לוג הנראות (כפונקציה של θ)

$$LL(\theta : D) = \sum_j LL(\theta : D^j) = \sum_j logP(X^j \mid \theta) = \sum_j log \sum_{S^j} P(X^j, S^j \mid \theta)$$

$$= \sum_j log \sum_{S^j} Q(S^j) \cdot \left[\frac{P(X^j, S^j \mid \theta)}{Q(S^j)} \right] \triangleq \sum_j log \mathbb{E}_Q \frac{P(X^j, S^j \mid \theta)}{Q(S^j)}$$

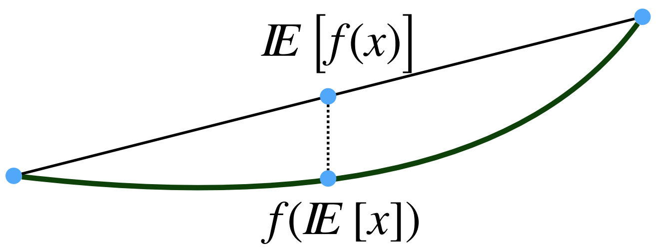
$$\overset{\text{Jensen}}{\geq} \sum_j \mathbb{E}_Q \log \frac{P(X^j, S^j \mid \theta)}{Q(S^j)} = \sum_j \mathbb{E}_Q \left[LL(\theta) \right] - \sum_j \sum_{S^j} Q(S^j) \log Q(S^j)$$

$$= \sum_j \mathbb{E}_Q \left[LL(\theta) \right] + \sum_j H(Q) \geq \sum_j \mathbb{E}_Q \left[LL(\theta) \right]$$

- כלומר סכום התוחלות לפי Q חוסם את לוג הניראות מלמטה

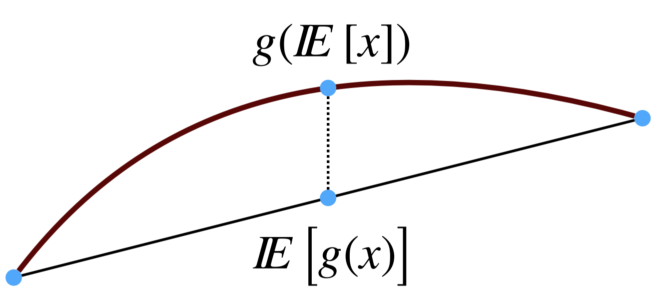
אי–שיוויון ינסן

- אי–שוויון ינסן: ממוצע ערכי פונקציה קמורה גדול/שווה לערך הפונקציה בממוצע הנקודות.



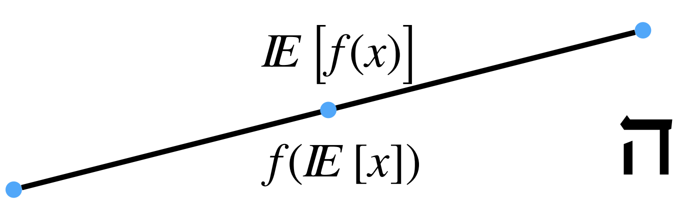
$$\mathbb{E}\left[f(x)\right] \geq f(\mathbb{E}\left[x\right]) \qquad f''(x) \geq 0$$

- ולחלופין, ממוצע ערכי פונקציה קעורה קטן/שווה לערך הפונקציה בממוצע הנקודות.



$$\mathbb{E}\left[g(x)\right] \leq g(\mathbb{E}\left[x\right]) \qquad g''(x) \leq 0$$

- פונקצית הלוג קעורה $\frac{d^2 \log x}{dx^2} = \frac{-1}{x^2} \leq 0$
- ולפיכך $\mathbb{E}\left[\log(x)\right] \leq \log(\mathbb{E}\left[x\right])$



- כמובן שיתקיים שיוויון אם f לינארית או קבועה

הבסיס התיאורטי ל-EM

- נחזור לחסם לוג הניראות בעזרת Q

$$LL(\theta) = \overset{\text{Jensen}}{\geq} \sum_j \mathbb{E}_Q\left[LL(\theta)\right] + \sum_j H(Q)$$

- איזה Q נבחר?
- נחפש Q עבורה $\dfrac{P(X^j, S^j \mid \theta)}{Q(S^j)}$ קבוע. כלומר $Q(S^j) \propto P(X^j, S^j \mid \theta)$
- למשל הפוסטריור: $P(X^j, S^j \mid \theta) = P(S^j \mid X^j, \theta) \cdot P(X^j \mid \theta)$
- ואז יתקיים: $LL(\theta) = \mathbb{E}_Q\left[LL(\theta)\right] + H(Q) \triangleq F_D\left[\theta, Q\right]$

$\mathrm{E_Q[LL]}$ נוחה למיקסום, וחוסמת את לוג הנראות מלמטה

הבסיס התיאורטי ל-EM

- נסכם.
- מדי צעד נעדכן את Q להיות **הפוסטריור** לפי θ הנוכחית

$$\text{E-step} : Q(S^j) = P(S^j \mid X^j, \theta)$$

- ואת θ למקסם את $\mathrm{E}_{\mathrm{Q}}[\mathrm{LL}]$ ובתוך כך לוג הניראות לפי θ

$$\text{M-step} : \tilde{\theta} = \arg\max_{\theta} \sum_j \mathbb{E}_Q \log P(X^j \mid \theta)$$

- בד"כ EM יתכנס תוך 5–10 איטרציות למקסימום לוקאלי