פיתרון חלקי מועד ב' – חורף 2020/21

פיתרון שאלה 1 [10 נק']

א. [5 נק'] נתון שבחלק מהעלים ישנן דוגמאות אימון ממספר מחלקות שונות.

א.1. מודל דטרמיניסטי:

דוגמה x תהיה שייכת למחלקה עם מס' דוגמאות האימון הגדול ביותר בעלה

א.2. מודל הסתברותי:

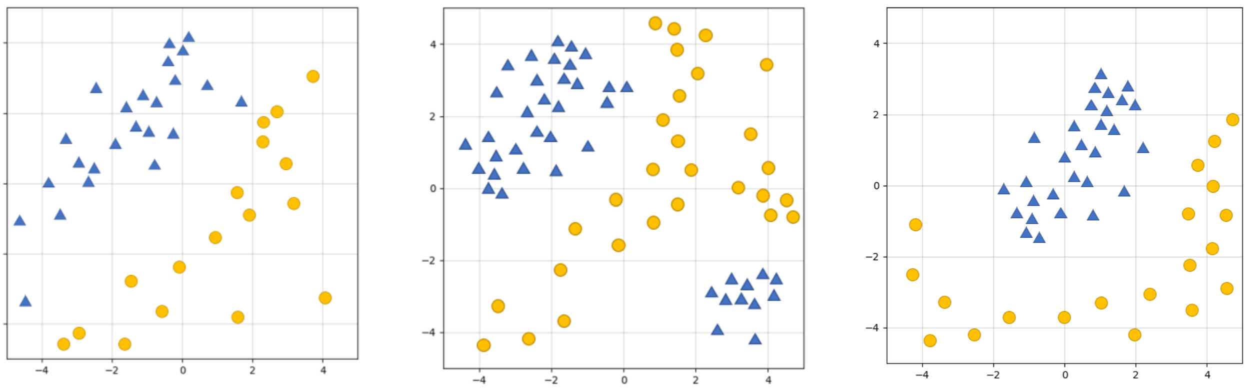
נעריך את התפלגות המחלקות בעלה ע"י מספר דוגמאות האימון הממופות לאותו עלה, דהיינו:

$$D_i = \frac{\left|\{x_j \mid j = i\}_{j=1..m}\right|}{m}$$

ב. [5 נק'] התופעה היא overfit, העץ מתאים את עצמו לסט האימון בצורה מושלמת אך בעל שגיאת הכללה גבוהה. ניתן לבצע pruning לענפי העץ על פי קריטריון מספר דוגמאות אימון מינימלי בעלה או לחילופין להגביל בצורה קשיחה את עומק העץ.

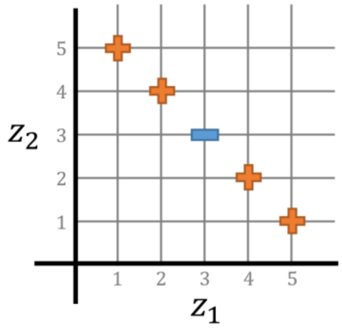
פיתרון שאלה 2 [10 נק']

A B C



מודלים (i,iii,iv)   מודלים (i,iii,iv)   מודלים (i,ii,iii)

שאלה 3 [20 נק']



א. [5 נק'] נציע את $h_1 = [z_1 < 6]$.

ב. [5 נק'] לדוגמה המסומנת $\boldsymbol{x}_3$.

ג. [5 נק'] ההתפלגות היא: $D_1^{(2)} = D_2^{(2)} = D_4^{(2)} = D_5^{(2)} = 1/8,\ \ D_3^{(2)} = 1/2$.

ד. [5 נק'] נציע את $h_2 = [z_1 < 2.5]$.

שאלה 4 [15 נק']

א. [5 נק'] הפיתוח:

$$\ln\big(p(\boldsymbol{w}|b_1, \dots, b_d)\big) = \ln\left(\prod_j p\big(w_j|b_j\big)\right) = \sum_j \ln\left(p\big(w_j|b_j\big)\right) = \sum_j \ln\left(\frac{1}{2b_j}\exp\left\{-\frac{|w_j|}{b_j}\right\}\right)$$
$$= \sum_j \left[\ln\left(\frac{1}{2b_j}\right) - \frac{|w_j|}{b_j}\right]$$

ב. [10 נק'] הוכחה:

$$\widehat{\boldsymbol{w}}_{\boldsymbol{\lambda}}^{\mathrm{AL}} = \widehat{\boldsymbol{w}}_{\boldsymbol{b}}^{\mathrm{MAP}} \triangleq \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmax}}\, p(\boldsymbol{w}|\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}, b_1, \dots, b_d)$$

$$\widehat{\boldsymbol{w}}_{\boldsymbol{b}}^{\mathrm{MAP}} \triangleq \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmax}}\, p(\boldsymbol{w}|\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}, b_1, \dots, b_d) = \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmax}} \ln p(\boldsymbol{w}|\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}, b_1, \dots, b_d)$$
$$= \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmax}} [\ln p(\boldsymbol{w}|b_1, \dots, b_d) + \ln p(\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}\,|\boldsymbol{w})]$$
$$= \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmax}} \left[\sum_j \left[\ln\left(\frac{1}{2b_j}\right) - \frac{|w_j|}{b_j}\right] - \frac{m}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{m} (\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle - y_i)^2\right]$$
$$= \underset{\boldsymbol{w}\in\mathbb{R}^{\mathrm{d}}}{\operatorname{argmin}} \left[\sum_{i=1}^{m} (\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle - y_i)^2 + 2\sum_j \frac{|w_j|}{b_j}\right]$$

ויש לבחור $b_j = 2/\lambda_j$.

פיתרון שאלה 5 [20 נק']

א. [1 נק'] לטובת הפתרון כאן, נניח ש $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$

ב. [6 נק'] נחשב:

$$LL_{p_1} = \log \prod_{i=1}^{4} \binom{10}{x_i} p_1^{x_i} (1-p_1)^{10-x_i} = \sum_{i=1}^{4} \log \binom{10}{x_i} + \sum_{i=1}^{4} \mathrm{x_i} \cdot \log p_1 + \sum_{i=1}^{4} (10-x_i) \log(1-p_1)$$

כעת, על מנת למצוא את נק' ה maximum likelihood, נגזור ונשווה ל 0.

$$\frac{\partial LL_{p_1}}{\partial p_1} = \sum_{i=1}^{4} \frac{\mathrm{x_i}}{p_1} - \sum_{i=1}^{4} \frac{10-x_i}{1-p_1} = 0 \ \rightarrow \ \ p_1 = \frac{\sum_{i=1}^{4} x_i}{10 \cdot n} = 0.25$$

ג. [6 נק'] נניח ש $\lambda = 0$ ולכן $p_1^{(0)} = 0.5,\ p_2^{(0)} = \frac{1}{3}$. מהנתונים ניתן להסיק ש $\Pr[z_i = 1] = 0.4, \Pr[z_i = 2] = 0.6$.

נשתמש בנוסחה לחישוב $Q_{i,j}$ ונרשום:

$$Q_{3,1}^{(1)} = \frac{\Pr[x_3, z_3 = 1]}{\sum_{j'=1,2} \Pr\,[x_3, z_3 = j'; \Theta^{(0)}]}$$

נחשב: $$\Pr[x_3, z_3 = 1] = \Pr\left[z_3 = 1; \Theta^{(0)}\right] \Pr[x_3 | z_3 = 1; \Theta^{(0)}] = 0.4 \cdot \binom{10}{x_3} p_1^{x_3} (1-p_1)^{10-x_3} = 0.046875$$

בדרך דומה נחשב $\Pr[x_3, z_3 = 2] = 0.1561$.

מכאן נציב בנוסחה לחישוב $Q_{i,j}$:

$$Q_{3,1}^{(1)} = 0.231, \qquad Q_{3,2}^{(1)} = 1 - Q_{3,1}^{(1)} = 0.76855$$

ד. [7 נק'] ראשית נבטא במפורש את F:

$$F(Q;\Theta) = \sum_i \sum_j Q_{i,j}^{(1)} \left(\log(\Pr[z_i = j]) + \log(\Pr[x_i \mid z_i = j])\right) =$$

$$= \sum_i Q_{i,1}^{(1)} \left(const_1 + \log\left(\binom{10}{x_i} p_1^{x_i} (1-p_1)^{10-x_i}\right)\right) + \sum_i Q_{i,2}^{(1)} \left(const_2 + \log\left(\binom{10}{x_i} p_2^{x_i} (1-p_2)^{10-x_i}\right)\right)$$

כעת נגזור את הפונקציה לפי p_1 ונשווה ל 0, על מנת למצוא את $p_1^{(1)}$:

$$\frac{\partial F(Q;\Theta)}{\partial p_1} = 0$$

$$\frac{\partial}{\partial p_1}\left(\sum_i Q_{i,1}^{(1)} \left(const_1 + \log\left(\binom{10}{x_i} p_1^{x_i} (1-p_1)^{10-x_i}\right)\right)\right) = 0$$

$$\frac{\partial}{\partial p_1}\left(\sum_i Q_{i,1}^{(1)} \left(const_1 + \log\left(\binom{10}{x_i}\right) + x_i \cdot \log(p_1) + (10 - x_i) \cdot \log(1-p_1)\right)\right) = 0$$

$$\sum_i Q_{i,1}^{(1)} \left(\frac{x_i}{p_1} - \frac{(10-x_i)}{1-p_1}\right) = 0$$

$$p_1^{(1)} = \frac{\sum_i Q_{i,1} \cdot x_i}{10 \cdot \sum_i Q_{i,1}} = 0.2716$$

שאלה 6 – שאלות קצרות [25 נק']

ו. [3 נק'] לכל אחת מהטענות הבאות, כתבו במחברת התשובות האם היא נכונה או לא נכונה (אין צורך להסביר).

a. LDA (Latent Discriminative Analysis) היא שיטת סיווג מונחית (supervised) באמצעות מפריד לינארי.

נכונה.

b. הזמן שלוקח לבצע איטרציית backpropagation אחת (על מעבד יחיד, ללא מִקבול) באימון של רשת עמוקה הוא לינארי בגודל הרשת. שימו לב: הכוונה ב"גודל הרשת" היא למספר הנוירונים ועוד מספר הקשתות. ניתן להניח שחישוב הנגזרת של פונקציית אקטיבציה לוקח זמן קבוע כלומר $O(1)$.

נכונה.

c. כש"מגדלים" עצי החלטה (Decision Trees), כדאי ליצור צמתים עם אנטרופיה (entropy) גבוהה.

לא נכונה.