שיטות כריית נתונים ובינה עסקית – 096411

מבחן סיום – מועד א'

מרצים: דר' דוד עזריאל, דר' תמיר חזן

מתרגל: ליאון ענבי

25 ביולי 2016

ת.ז.:________________

<u>הוראות – נא לקרוא בעיון רב</u>

- משך הבחינה 3 שעות.
- חומר עזר מותר לבחינה הינו מחשבון וכל חומר כתוב.
- <u>אין להפריד</u> אף דף מטופס הבחינה.
- במבחן זה ארבע שאלות ובכל שאלה מספר סעיפים. יש לבחור שלוש מתוכן ולענות עליהן במלואן. משקלה של כל שאלה הינו 33 נקודות כאשר לציון הסופי תתווסף נקודה אחת נוספת.
- <u>שימי לב:</u> תיבדקנה רק שלוש השאלות הראשונות לפי סדר הופעתן במחברת הבחינה. אין טעם לפתור יותר משלוש שאלות מכיוון שהשאלה הרביעית לא תיבדק.
- המבחן מנוסח בלשון נקבה אך מתייחס לשני המינים
- בסיום המבחן יש למסור את טופס הבחינה.
- בהצלחה!!!

שאלה 1 (33 נק')

נתונות n תצפיות חד מימדיות $X_1, \dots, X_n$ אשר מגיעות מעירוב של שתי קבוצות. התצפיות בכל כל קבוצה מתפלגות מעריכית. כל תצפית מגיעה מהקבוצה הראשונה בהסתברות π.

<u>תזכורת:</u> פונקציית הצפיפות של משתנה מקרי מעריכי עם פרמטר θ היא $\theta e^{-\theta x}$.

א. (10 נק') הראי כי פונקציית *log-likelihood* של הנתונים הינה:

$$l(X|\theta_1, \theta_2, \pi) = \sum_{i=1}^{n} log\left(\pi\theta_1 e^{-\theta_1 x_i} + (1-\pi)\theta_2 e^{-\theta_2 x_i}\right)$$

<u>פתרון</u>

נגדיר את Δ להיות משתנה מקרי אשר מקבל את הערכים $\{1,2\}$ ומייצג את הקבוצה ממנה הגיעה התצפית. הצפיפות המותנית:

$$f(x|\Delta = 1) = \theta_1 e^{-\theta_1 x_i}, \ \ f(x|\Delta = 2) = \theta_2 e^{-\theta_2 x_i}$$

מנוסחת ההסתברות השלמה

$$f(x) = p(\Delta = 1) f(x|\Delta = 1) + p(\Delta = 2) f(x|\Delta = 2) = \pi\theta_1 e^{-\theta_1 x} + (1-\pi)\theta_2 e^{-\theta_2 x}$$

פונקציית הנראות של הנתונים

$$L(X, \theta_1, \theta_2, \pi) = \prod_{i=1}^{n} \pi\theta_1 e^{-\theta_1 x_i} + (1-\pi)\theta_2 e^{-\theta_2 x_i}$$

פונקציית *log-likelihood*

$$l(X|\theta_1, \theta_2, \pi) = \log\big(L(X, \theta_1, \theta_2, \pi)\big) = \log\left(\prod_{i=1}^{n} \pi\theta_1 e^{-\theta_1 x_i} + (1-\pi)\theta_2 e^{-\theta_2 x_i}\right)$$

$$= \sum_{i=1}^{n} log\left(\pi\theta_1 e^{-\theta_1 x_i} + (1-\pi)\theta_2 e^{-\theta_2 x_i}\right)$$

ב. (15 נק') תארי אלגוריתם *EM* לאמידת הפרמטרים π, θ_1, θ_2.

<u>תזכורת:</u> עבור $Y_1, \dots, Y_m \sim \exp(\theta)$ אומד נראות מרבית ל- θ הוא $\hat{\theta} = \frac{1}{\frac{1}{m}\sum_{i=1}^{m} Y_i}$

<u>פתרון</u>

נתחיל אם ערכים התחלתיים כלשהם עבור הפרמטרים $\hat{\pi}, \hat{\theta}_1, \hat{\theta}_2$ ונבצע איטרציות של הצעדים E ואחריו M עד להתכנסות.

צעד E:

$$\hat{P}(\Delta_i=1)=P_{\hat{\pi},\hat{\theta}_1,\hat{\theta}_2}(\Delta_i=1|X_i)=\frac{P(X_i,\Delta_i=1)}{P(X_i)}$$

$$=\frac{P(\Delta_i=1)P_{\hat{\pi},\hat{\theta}_1,\hat{\theta}_2}(X_i|\Delta_i=1)}{P(\Delta_i=1)P_{\hat{\pi},\hat{\theta}_1,\hat{\theta}_2}(X_i|\Delta_i=1)+P(\Delta_i=2)P_{\hat{\pi},\hat{\theta}_1,\hat{\theta}_2}(X_i|\Delta_i=2)}$$

$$=\frac{\hat{\pi}\hat{\theta}_1e^{-\hat{\theta}_1x_i}}{\hat{\pi}\hat{\theta}_1e^{-\hat{\theta}_1x_i}+(1-\hat{\pi})\hat{\theta}_2e^{-\hat{\theta}_2x_i}}$$

צעד M:

אמדי נראות מירבית:

$$\hat{\pi}=\frac{\sum_{i=1}^{n}\hat{P}(\Delta_i=1)}{n},\qquad \hat{\theta}_1=\frac{1}{\frac{\sum_{i=1}^{n}\hat{P}(\Delta_i=1)\,x_i}{\sum_{i=1}^{n}\hat{P}(\Delta_i=1)}},$$

$$\hat{\theta}_2=\frac{1}{\frac{\sum_{i=1}^{n}\hat{P}(\Delta_i=2)\,x_i}{\sum_{i=1}^{n}\hat{P}(\Delta_i=2)}}=\frac{1}{\frac{\sum_{i=1}^{n}\left(1-\hat{P}(\Delta_i=1)\right)x_i}{\sum_{i=1}^{n}\left(1-\hat{P}(\Delta_i=1)\right)}}$$

נניח כי התקבלו בסעיפים הקודמים האומדים הבאים: $\hat{\pi}=0.6,\hat{\theta}_1=\frac{1}{3},\hat{\theta}_2=\frac{1}{4}$.

ג. (8 נק') מאיזו קבוצה סביר יותר לקבל תצפית שערכה $X=5$? הסבירי תשובתך.

פתרון

$$\hat{P}(\Delta_i=1)=P_{\hat{\pi},\hat{\theta}_1,\hat{\theta}_2}(\Delta_i=1|X=5)=\frac{\hat{\pi}\hat{\theta}_1e^{-\hat{\theta}_1x_i}}{\hat{\pi}\hat{\theta}_1e^{-\hat{\theta}_1x_i}+(1-\hat{\pi})\hat{\theta}_2e^{-\hat{\theta}_2x_i}}$$

$$=\frac{0.6\frac{1}{3}e^{-\frac{1}{3}5}}{0.6\frac{1}{3}e^{-\frac{1}{3}5}+0.4\frac{1}{4}e^{-\frac{1}{4}5}}=\frac{0.0377}{0.0377+0.0286}=0.56$$

סביר יותר שהתצפית הגיעה מקבוצה 1.

שאלה 2 (33 נק')

נתון מודל רגרסיה פשוטה: $\mathrm{Y}_i = \beta_0 + \beta_1 X_i + \varepsilon_i \;\; i = 1, \dots, n$ כאשר מתקיים $\sum_{i=1}^{n} X_i = 0$

א. (11 נק') הראי שאומד *RIDGE* ל-$\widehat{\beta}_1$ עבור פרמטר λ מסוים הוא:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2 + \lambda}$$

פתרון

אומד *RIDGE* הינו:

$$\hat{\beta} = argmin_{(\beta_0,\beta_1)} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 + \lambda\beta_1^2 = argmin_{(\beta_0,\beta_1)} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 + \lambda\beta_1^2$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \widehat{\beta}_1 X_i\right)^2 + \lambda \widehat{\beta}_1^2 = -2 \sum_{i=1}^{n} X_i \left(Y_i - \hat{\beta}_0 - \widehat{\beta}_1 X_i\right) + 2\lambda \widehat{\beta}_1 = 0$$

$$\sum_{i=1}^{n} X_i \left(Y_i - \hat{\beta}_0 - \widehat{\beta}_1 X_i\right) = \lambda \widehat{\beta}_1$$

$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \widehat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \lambda \widehat{\beta}_1$$

$$\sum_{i=1}^{n} X_i Y_i = \widehat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 + \lambda \right)$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2 + \lambda}$$

ב. (11 נק') הראי שהתוחלת והשונות של האומד מסעיף א' הינן:

$$E\left(\hat{\beta}_1\right) = \beta_1 \frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} X_i^2 + \lambda} \qquad\qquad Var\left(\hat{\beta}_1\right) = \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{\left(\sum_{i=1}^{n} X_i^2 + \lambda\right)^2}$$

פתרון

$$E(\hat{\beta}_1) = E\left(\frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2 + \lambda}\right) = \frac{1}{\sum_{i=1}^n X_i^2 + \lambda} E\left(\sum_{i=1}^n Y_i X_i\right) = \frac{1}{\sum_{i=1}^n X_i^2 + \lambda}\sum_{i=1}^n E(Y_i X_i)$$

$$= \frac{\sum_{i=1}^n X_i E(\beta_0 + \beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^n X_i^2 + \lambda} = \frac{\sum_{i=1}^n X_i\big(E(\beta_0) + E(\beta_1 X_i) + E(\varepsilon_i)\big)}{\sum_{i=1}^n X_i^2 + \lambda}$$

$$= \frac{\sum_{i=1}^n X_i(\beta_0 + \beta_1 X_i)}{\sum_{i=1}^n X_i^2 + \lambda} = \frac{\sum_{i=1}^n X_i\beta_0 + \sum_{i=1}^n X_i(\beta_1 X_i)}{\sum_{i=1}^n X_i^2 + \lambda} = \frac{\beta_1\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2 + \lambda}$$

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2 + \lambda}\right) = \frac{1}{(\sum_{i=1}^n X_i^2 + \lambda)^2} Var\left(\sum_{i=1}^n Y_i X_i\right) = \frac{\sum_{i=1}^n Var(Y_i X_i)}{(\sum_{i=1}^n X_i^2 + \lambda)^2}$$

$$= \frac{\sum_{i=1}^n X_i^2 Var(Y_i)}{(\sum_{i=1}^n X_i^2 + \lambda)^2} = \frac{\sigma^2\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n X_i^2 + \lambda)^2}$$

ג. (11 נק') האם ההטיה בריבוע של האומד מסעיף א', $\left(E[\hat{\beta}_1 - \beta_1]\right)^2$ עולה או יורדת ב-λ? האם השונות של האומד עולה או יורדת ב-λ? הסבירי את התוצאה תוך התייחסות למטרה של השימוש ב-*RIDGE*.

<u>פתרון</u>

ההטיה בריבוע של האומד מסעיף א':

$$\left(E[\hat{\beta}_1 - \beta_1]\right)^2 = \left(\frac{\beta_1\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2 + \lambda} - \beta_1\right)^2 = \left(\beta_1\left(\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2 + \lambda} - 1\right)\right)^2 = \beta_1^2\left(\frac{-\lambda}{\sum_{i=1}^n X_i^2 + \lambda}\right)^2$$

$$= \beta_1^2\left(\frac{1}{\frac{\sum_{i=1}^n X_i^2}{\lambda} + 1}\right)^2$$

בביטוי הנ"ל המכנה יורד ב-λ ומכאן שהביטוי כולו עולה ב-λ. ניתן לראות שכאשר $\lambda = 0$ מתקבל אמד חסר הטיה (רגרסיה לינארית רגילה) וכאשר $\lambda \to \infty$ נקבל שההטיה של האומד הינה β_1 מכיוון ש-$\hat{\beta}_1 = 0$.

בשונות של האומד מסעיף א', המכנה עולה ב-λ ומכאן שהביטוי כולו יורד ב-λ.

שתי התוצאות מסתדרות היטב עם המטרה של *RIDGE regularization* שהיא להביא למודל בעל שונות נמוכה יותר. עושים זאת ע"י הגבלת ערכי האומד סביב אפס ובכך למעשה כופים שגיאה על המודל בכיוון שבו המקדמים שווים לאפס.

שאלה 3 (33 נק')

נתונות m תצפיות $x_1, \dots, x_m$ ממימד $p > 2$ אשר מקיימות

$$\frac{1}{m}\sum_{i=1}^{m} x_i = 0$$

נסמן ב-X את המטריצה שמכילה את התצפיות בעמודות ואת מטריצת השונויות של התצפיות ב-$XX^T = \Sigma$.

עבור כל וקטור u נגדיר את $y = u^T x$ להיות הטרנספורמציה של x באמצעות u.

נחפש את וקטור היחידה u ($u^T u = 1$) אשר ממקסם את השונות של התצפיות $y_1, \dots y_m$:

$$u_1 = argmax_u \left\{ \sum_{i=1}^{m} (y_i - \bar{y})^2 \right\}, \qquad s.t.\ \ u^T u = 1$$

א. (5 נק') הסבירי לשם מה נחוץ האילוץ ש-u הינו וקטור יחידה.

<u>פתרון</u>

אם לא נגביל את וקטור u להיות וקטור היחידה, עבור כל וקטור u נוכל לייצר וקטור $v = Cu, C > 1$ שעבורו:

$$\sum_{i=1}^{m} \left(y^{(v)}{}_i - \bar{y}\right)^2 = \sum_{i=1}^{m} y^{(v)}{}_i{}^2 = \sum_{i=1}^{m} (v^T x_i)^2 = \sum_{i=1}^{m} (C u^T x_i)^2 = C^2 \sum_{i=1}^{m} (u^T x_i)^2 > \sum_{i=1}^{m} (u^T x_i)^2$$

בכך למעשה הראנו כי לא נוכל למצוא וקטור שממקסם את הביטוי בלי ההגבלה על וקטור u.

באופן שקול ניתן לבטא את האילוץ הנ"ל ע"י שימוש בכופלי לגרנז' וניסוח בעיית האופטימיזציה הבאה:

$$u_1 = argmax_u \left\{ \sum_{i=1}^{m} (y_i - \bar{y})^2 - \lambda (u^T u - 1) \right\}$$

ב. (15 נק') הראי כי u_1 הינו וקטור עצמי של מטריצת השונויות Σ

תזכורת: וקטור עצמי v למטריצה A עם ערך עצמי λ שמתאים לו מקיימים:

$$Av = \lambda v$$

<u>פתרון</u>

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{m} (y_i - \bar{y})^2 - \lambda (u^T u - 1) = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{m} (y_i)^2 - \lambda (u^T u - 1) = \frac{\partial}{\partial \beta_1} \left(u^T \Sigma u - \lambda (u^T u - 1) \right)$$

$$= 2\Sigma \mathrm{u} - 2\lambda u = 0$$

$$\Sigma \mathrm{u} = \lambda u$$

ג. (5 נק') הראי כי מבין הוקטורים העצמיים של Σ, הוקטור העצמי u_1 אשר ממקסם את השונות של Y יהיה הוקטור העצמי אשר מתאים לערך העצמי המקסימלי.

רמז: זכרי כי u_1 הינו: (a) וקטור עצמי של Σ, (b) וקטור יחידה.

<u>פתרון</u>

נזכור כי נרצה למקסם את השונות של y $u_k^T \Sigma u_k = u_k^T \lambda_k u_k = \lambda_k u_k^T u_k = \lambda_k$ ומכאן שלבחור את λ_k המקסימלי ייתן את השונות המקסימלית.

ד. (8 נק') נתונות התצפיות הבאות: $(-1,1),(0,0),(1,-1)$

מצאי את הוקטור u_1 עבורן, הציגי את ערכי y_1, y_2, y_3 ואת שגיאת השחזור של התצפיות.

<u>פתרון</u>

שלו הנקודות ממורכזות סביב אפס ונמצאות על קו ישר. שני וקטורים אפשריים: $v_1 = (1,-1), v_2 = (-1,1)$.

ננרמל את הוקטורים כדי שיהיו וקטורי יחידה $u_1 = (\sqrt{2}, -\sqrt{2}), u_2 = (-\sqrt{2}, \sqrt{2})$

לפי u_1:

$$y_1 = u_1^T x_1 = \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix} (-1,1) = -2\sqrt{2}$$

$$y_2 = u_1^T x_2 = \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix} (0,0) = 0$$

$$y_3 = u_1^T x_3 = \begin{pmatrix} \sqrt{2} \\ -\sqrt{2} \end{pmatrix} (1,-1) = 2\sqrt{2}$$

מכיוון שהתצפיות נמצאות על קו ישר, השחזור שלהם לאחר הטלה באמצעות u_1 מחזירם אותן למקומן המקורי ושגיאת השחזור הינה 0.

שאלה 4 (33 נק')

בידינו m תצפיות $(x_1, y_1), \dots, (x_m, y_m)$ כאשר $y \in \{1, \dots, k\}$, לכל (x, y) נגדיר וקטור $\phi(x, y)$ ונתייחס לפונקציית ההסתברות הבאה:

$$p(y|x, w) = \frac{e^{w^T\phi(x,y)}}{\sum_{y'=1}^{k} e^{w^T\phi(x,y')}}$$

במקרה הבינארי $y \in \{-1, +1\}$ הגדרנו:

$$p(y = +1|w, x) = \frac{e^{w^T x}}{1 + e^{w^T x}}$$

א. (5 נק') הראי כי המקרה הבינארי הינו מקרה פרטי של המקרה ובו k מחלקות. כלומר, הראי כי קיים $\phi(x, y)$ עבורו פונקציות ההסתברות שוות.

פתרון

נגדיר את $\phi(x, y)$ במפורש:

$$\phi(x, y) = \begin{Bmatrix} x, y = 1 \\ 0, y = 0 \end{Bmatrix}$$

ונקבל:

$$p(y = +1|x, w) = \frac{e^{w^T\phi(x,1)}}{e^{w^T\phi(x,1)} + e^{w^T\phi(x,0)}} = \frac{e^{w^T x}}{e^{w^T x} + e^{w^T 0}} = \frac{e^{w^T x}}{e^{w^T x} + 1}$$

אמד נראות מרבית יהיה:

$$w^* = argmax_w \left\{ \sum_{i=1}^{m} \log\big(p(y_i|x_i, w)\big) \right\}$$

ב. (13 נק') עבור המקרה הבינארי, הראי כי מתקיים:

$$\sum_{i=1}^{m} p(y = +1|x_i, w^*) x_i = \sum_{i; y_i = +1} x_i$$

שימי לב כי x, w שניהם וקטורים.

פתרון

נראה כי עבור כל רכיב $r = 1, \dots p$ מתקיים השוויון:

$$\sum_{i=1}^{m} p(y = +1|x_i, w^*) x_{i,r} = \sum_{i; y_i = +1} x_{i,r}$$

נגזור את פונקציית הנראות לפי הרכיב w_r

$$\frac{\partial}{\partial w_r}\sum_{i=1}^{m}\log\big(p(y_i|x_i,w)\big)=\frac{\partial}{\partial w_r}\sum_{i;y_i=+1}\log\left(\frac{e^{w^Tx_i}}{1+e^{w^Tx_i}}\right)+\sum_{i;y_i=-1}\log\left(\frac{1}{1+e^{w^Tx_i}}\right)$$

$$=\frac{\partial}{\partial w_r}\sum_{i;y_i=+1}\left(w^Tx_i-\log\big(1+e^{w^Tx_i}\big)\right)-\sum_{i;y_i=-1}\log\big(1+e^{w^Tx_i}\big)$$

$$=\frac{\partial}{\partial w_r}\sum_{i;y_i=+1}w^Tx_i-\sum_{i=1}^{m}\log\big(1+e^{w^Tx_i}\big)=$$

$$\sum_{i;y_i=+1}x_{i,r}-\sum_{i=i}^{m}\frac{e^{w^Tx_i}}{1+e^{w^Tx_i}}x_{i,r}=\sum_{i;y_i=+1}x_{i,r}-\sum_{i=1}^{m}p(y=+1|x_i,w)x_{i,r}$$

נשווה לאפס למציאת נקודת קיצון:

$$\sum_{i;y_i=+1}x_{i,r}-\sum_{i=1}^{m}p(y=+1|x_i,w^*)x_{i,r}=0$$

$$\sum_{i=1}^{m}p(y=+1|x_i,w^*)x_{i,r}=\sum_{i;y_i=+1}x_{i,r}$$

כעת נניח כי שלושה סטודנטים מימשו שלושה מסווגים f_1,f_2,f_3 כאשר כל מסווג הינו פונקציה $f\big((X_{train}Y_{train}),x_{new}\big)\in\{1,\dots,m\}$ אשר מקבלת תצפיות אימון $(X_{train}Y_{train})$ ומחזירה סיווג (מחלקה) עבור התצפית החדשה x_{new}.

על מנת להשוות בין המסווגים ולקבוע מי מהם בעל ביצועים טובים יותר חילקנו את m התצפיות לשתי קבוצות G_1,G_2 וחישבנו את הערכים הבאים לכל מסווג:

$$E_{1,1}=\frac{100}{|G_1|}\sum_{(x_i,y_i)\in G_1}I\{f(G_1,x_i)=y_i\}$$

$$E_{1,2}=\frac{100}{|G_2|}\sum_{(x_i,y_i)\in G_2}I\{f(G_1,x_i)=y_i\}$$

ג. (5 נק') מהם שני הערכים $E_{1,1},E_{1,2}$? הסבירי תשובתך בקצרה.

<u>פתרון</u>

$E_{1,1}$ הינם אחוזי ההצלחה של מסווג הנבנה על הקבוצה G_1 על נתוני אימון. אנחנו מאמנים מסווג באמצעות התצפיות שבקבוצה G_1 ומחשבים את פרופורציית הסיווגים המוצלחים בקבוצה G_1.

$E_{1,2}$ הינם אחוזי ההצלחה של מסווג הנבנה על הקבוצה G_1 על נתוני מבחן. אנחנו מאמנים מסווג באמצעות התצפיות שבקבוצה G_1 ומחשבים את פרופורציית הסיווגים המוצלחים בקבוצה G_2 שתצפיותיה לא נכללו בבניית המסווג.

<u>מי שענה כאילו מדובר בשגיאות אימון ומבחן יקבל את מרבית הניקוד בתנאי שההסבר נכון עד כדי הסימן באינדיקטור.</u>

ד. (10 נק') התקבלה התוצאה הבאה עבור שלושת המסווגים:

	f_1	f_2	f_3
$E_{1,1}$	*10*	10	*12*
$E_{1,2}$	*15*	50	*12*

האם ניתן לומר כי המסווג f_3 עדיף על סמך התוצאות? הציעי לפחות בדיקה נוספת אחת שתשפר את יכולת ההחלטה לגבי מי המסווג העדיף מבין השלושה ונמקי את תשובתך.

<u>פתרון</u>

על פניו נראה כי מסווג f_2 עדיף מכיוון שהוא בעל אחוזי הצלחה גבוהים יותר משני המסווגים האחרים על נתוני המבחן. עם זאת, מכיוון שביצענו **<u>חלוקה יחידה</u>** של התצפיות למדגם אימון ומדגם מבחן ומכיוון שלא צוינו גודל הקבוצות G_1, G_2 לא ניתן לקבוע זאת בביטחון.

על מנת לשפר את יכולת ההחלטה לגבי המסווג העדיף אבחר לבצע *10-fold CV* שבמהלכו אחלק את הנתונים לעשר קבוצות ועבור כל קבוצה אחשב את אחוזי הצלחה של הקבוצה באמצעות שלושת המסווגים תוך שימוש בתשע הקבוצות הנותרות בתור נתוני האימון. בדרך זה אקבל עבור כל מסווג 10 ערכים לאחוזי הצלחה. בשלב הבא אחשב ממוצע ל-10 הערכים עבור כל מסווג לקבלת אחוזי הצלחה ממוצעים לכל מסווג ואשווה בין הנ"ל לבחירת המסווג העדיף.

השימוש בחלוקות השונות עוזר להימנע מתוצאות מקריות שנובעות מחלוקה מקרית אחת של הנתונים לקבוצת נתוני אימון וקבוצת נתוני מבחן.

<u>מי שענה כאילו מדובר בשגיאות אימון ומבחן יקבל ניקוד בהנחה שהתשובה מתאימה לתשובה לסעיף הקודם וכי ההסברים נכונים..</u>