<u>תרגיל 1 בלמידת מכונה שימושית</u>

מגישה: דנה אבירן

<u>מקרה 1 – 2 נביאים, משחק יחיד</u>

2 נביאים עם הסתברויות 0.2, 0.4 לשגיאה בהתאמה, שנבחר ביניהם בשימוש במזעור סיכונים אמפירי (ERM) על פני Train Set של משחק יחיד. נחזור על הניסוי הזה 100 פעמים.

```
Scenario 1 Results:
Number of times best prophet selected: 59
Average test error of selected prophet: 0.28106999999999993
Average approximation error: 0.2
Average estimation error: 0.08200000000000003
```

<u>הסבר</u> - כאשר הנביא הראשון חזה תוצאה נכונה והנביא השני חזה תוצאה שגויה, הנביא הראשון נבחר בהתבסס על עקרון ERM. כאשר הנביא הראשון חזה תוצאה שגויה והנביא השני חזה תוצאה נכונה, הנביא השני נבחר. כאשר שני הנביאים מגיעים לאותה התוצאה (שניהם צודקים או שניהם טועים) האלגוריתם יבחר רנדומלית ביניהם בסיכוי שווה. תהליך שבירת השוויון האקראי מוסיף רובד של ריאליזם למנגנון הבחירה ומכיר בכך שבפועל, בחירה בין מודלים בעלי ביצועים דומים יכולה להיות מורכבת כיוון שלא יהיה לנו מידע על אחוז השגיאות האמיתי של המודלים.

<u>תשובות לשאלות</u>

1. השגיאה הממוצעת של הנביא הנבחר היא בערך 0.28.
   ערך זה מייצג את ההפרש הממוצע בין החיזויים של הנביא שנבחר על ה-Test set לבין התוצאות בפועל.
2. ב-59 מתוך 100 ניסויים, הנביא הראשון נבחר כנביא הכי טוב בהתבסס על עקרון מזעור סיכונים אמפירי (ERM). תוצאה זו הגיונית כי הסיכוי שלנו לבחור את הנביא הראשון בניסוי יחיד כזה הייתה הסיכוי שהתרחיש שבו הנביא הראשון חוזה נכון והנביא השני חוזה לא נכון או ששני הנביאים חוזים נכון וגם הנביא הראשון ייבחר אקראית בסיכוי 0.5. כלומר היה לנו סיכוי של 0.6[1] לבחור את הנביא הנכון.
3. חישוב שגיאות:
   a. mean approximation error = 0.2. ה-true risk של הנביא האופטימלי בסט הנביאים.
   b. mean estimation error - טעות ההערכה הממוצעת היא בערך 0.08. טעות ההערכה הממוצעת מחושבת על ידי סכום ההפרש המוחלט של ה-true risk של הנביא האופטימלי לבין ה-true risk של הנביא שנבחר בכל ניסוי, בחלוקה של מספר הניסויים. במקרה הנ"ל, טעות ההערכה הממוצעת בניסוי מסוים שווה ל-0 כאשר נבחר הנביא האופטימלי ושווה ל-0.2 כאשר נבחר הנביא השני, כי ההפרש הוא 0.2=0.4-0.2. לכן, כיוון שבקירוב יש לנו סיכוי של 0.4 לבחור בנביא השני (מתוך החישובים בסעיף 2), באמת היינו מצפים שחישוב ה-mean estimation error יצא 0.08 = (40/100)*0.2 .

<u>מקרה 2 – 2 נביאים, 10 משחקים</u>

2 נביאים עם הסתברויות 0.2, 0.4 לשגיאה בהתאמה, שנבחר ביניהם בשימוש ב-ERM על פני Train Set של עשרה משחקים. נחזור על הניסוי הזה 100 פעמים.

```
Scenario 2 Results:
Number of times best prophet selected: 82
Average test error of selected prophet: 0.23905
Average approximation error: 0.2
Average estimation error: 0.036000000000000001
```

<u>תשובות לשאלות</u>

1. השגיאה הממוצעת של הנביא הנבחר היא בערך 0.24, לעומת 0.28 בערך במקרה הקודם. הירידה בשגיאה נגרמה מכך שהנביא האופטימלי נבחר מספר רב יותר של פעמים (ראו הסבר בסעיף 2) במקרה זה מאשר במקרה הקודם, ולכן השגיאה הממוצעת כעת מושפעת באחוזים גבוהים יותר מתוצאות חיזוייו של הנביא האופטימלי, הנביא הראשון מאשר תוצאותיו של הנביא השני.
2. ב-82 מתוך 100 ניסויים הנביא הראשון נבחר כנביא הכי טוב, לעומת בחירתו 59 פעמים במקרה הקודם. ניתן לראות שבהתחשב בתוצאות של עשרה ניסויים במקום ניסוי יחיד, הסיכוי שלנו לבחור את הנביא הראשון גדל משמעותית. על פי עיקרון ERM, שיטת הבחירה בנביא הכי טוב לאחר עשרה משחקים היא בחירת הנביא בעל השגיאה הממוצעת הקטנה ביותר. לכן, בחירת הנביא השני תהיה נדירה יותר כי הסיכוי שלו להצליח לנבא רנדומלית נכון קטן כאשר הנביאים מתחרים על מספר גדול יותר של משחקים. ניתן להסביר זאת אינטואיטיבית באמצעות חוק המספרים

---

[1] (0.8 * 0.4) + (0.8 * 0.6 * 0.5) + (0.2 * 0.4 * 0.5) = 0.6 (הסתברות שלמה, נביאים בלתי-תלויים, בחירה רנדומלית (שוברת-שוויון אחידה)

הגדולים, שממנו נובע שהשגיאה הממוצעת על פני עשרה משחקים תתקרב לאחוז השגיאה האמיתי. כלומר, ככל שמספר הדוגמאות עולה כך קטן ה-generalization error, ההפרש המוחלט בין ה-empirical risk לבין ה-true risk. ואכן, התוצאות הללו עקביות עם אחוז השגיאה (true risk) 0.2 של הנביא הראשון, שכן היינו מצפים שהוא ייבחר בערך ב-80 אחוז מהניסויים.

3. חישוב שגיאות:
   a. Mean Approximation Error = 0.2, בהסבר זהה למקרה הקודם.
   b. Mean Estimation Error - טעות ההערכה הממוצעת היא בערך 0.036, לעומת 0.082. בגלל שהנביא האופטימלי נבחר פעמים רבות יותר טעות ההערכה המינימלית קטנה.

<u>מקרה 3 – 500 נביאים, עשרה משחקים</u>

נגריל 500 נביאים רנדומליים עם שגיאות שמתפלגות אחיד על [0,1]. נבחר ביניהם בשימוש ב-ERM על פני Train Set של עשרה משחקים. נחזור על הניסוי 100 פעמים.

```
Scenario 3 Results:
Average test error of selected prophet: 0.08951999999999995
Number of times best prophet selected: 0
Number of times prophet selected within epsilon: 0
Average approximation error: 0.00044603170617207866
Average estimation error: 0.08859550486697752
```

<u>תשובות לשאלות</u>

1. השגיאה הממוצעת של הנביא הנבחר היא בערך 0.09.
2. לא בחרנו את הנביא הכי טוב אפילו פעם אחת. הסבר אפשרי הוא שמספר המשחקים (גודל ה-train set) קטן משמעותית ממספר הנביאים ולכן סביר שנביאים בעלי true risk גבוה יותר מאשר הנביא האופטימלי יצליחו לנבא נכון באחוזים גבוהים או אפילו בכל המשחקים, כך שלפי עיקרון ERM נבחר בראשון מבין אלו שביצעו טוב, ללא התחשבות ב-true risk שלהם. בנוסף, מספר הנביאים הגדול גורם לכך שה-true risk של הנביא האופטימלי הוא ממש נמוך, כך שגם נביאים בעלי אחוזי שגיאה גדולים משמעותית יוכלו לנבא נכון ולכן ייבחרו.
3. לא בחרנו נביא שגרוע בפחות מאפסילון (0.01) אפילו פעם אחת. זהה להסבר בסעיף הקודם - סביר שנביאים בעלי true risk גבוה אף יותר מ-0.01 מאשר ה-true risk של הנביא האופטימלי יצליחו לנבא נכון באחוזים גבוהים או אפילו בכל המשחקים, כך שלפי עיקרון ERM נבחר בראשון מבין אלו שביצעו טוב.
4. חישוב השגיאות
   a. Mean Approximation Error - טעות הקירוב הממוצעת היא בערך 0.0004.
   b. Mean Estimation Error - טעות ההערכה הממוצעת היא בערך 0.088 (אכן גדול בהרבה מ-0.01).
5. אם ה-Error rates היו מתפלגים אחיד על [0, 0.5], במקום על [0,1], מה היה קורה לשגיאות? השגיאות היו קטנות. ראשית, ככל שטווח השגיאה האפשרי של הנביא הרנדומלי קטן יותר, כך מלכתחילה גובר הסיכוי לבחור נביאים שטווח אחוזי השגיאה שלהם יהיה קטן יותר. בנוסף, גובר הסיכוי לבחור נביאים שטווח אחוזי השגיאה שלהם קרוב יותר לטווח השגיאה של הנביא האופטימלי. לכן, נצפה שה-approximation error, estimation error וה-test error יהיו בעלי אחוזי שגיאה קטנים יותר בממוצע. הרצה בפועל של מקרה זה בו אחוזי השגיאה נבחרים מתוך הטווח [0,0.5] אכן מניבה אחוזי שגיאה קטנים יותר:

```
Average test error of selected prophet: 0.07097999999999995
Number of times best prophet selected: 0
Number of times prophet selected within epsilon: 0
Average approximation error: 0.00022301585308603933
Average estimation error: 0.0718527442736805
```

<u>מקרה 4 – 500 נביאים, אלף משחקים</u>

נגריל 500 נביאים רנדומליים עם שגיאות שמתפלגות אחיד על [0,1]. נבחר ביניהם בשימוש ב-ERM על פני Train Set של אלף משחקים. נחזור על הניסוי 100 פעמים.

<u>תוצאות</u>

```
Scenario 4 Results:
Average test error of selected prophet: 0.004710000000000003
Number of times best prophet selected: 25
Number of times prophet selected within epsilon: 100
Average approximation error: 0.003424889449009094
Average estimation error: 0.0009202354454096883
Average train generalization error: 0.0022910878940556214
Average test generalization error: 0.002800000000000002
```

<u>תשובות לשאלות</u>

1. השגיאה הממוצעת של הנביא הנבחר היא 0.0047, לעומת 0.09 בניסוי הקודם. ניתן לראות שהשגיאה הממוצעת קרובה מאוד ל-approximation error, כלומר במקרה הזה הנביאים שנבחרו ביצעו בצורה קרובה מאוד ל-true risk של הנביא האופטימלי. ניתן להסביר ירידה דרמטית זו בשגיאה הממוצעת על ידי התייחסות לגודל ה-Train Set. כלומר, בדומה לשינוי שראינו במקרה 2 לעומת מקרה 1, הסיכוי שלנו לבחור נביא עם true risk נמוך יותר בהתחשב בתוצאות של מספר גדול יותר של ניסויים גדל, כי מחוק המספרים הגדולים, השגיאה הממוצעת (empirical risk) על פני יותר משחקים תתקרב לאחוז השגיאה האמיתי, וכך יהיה ניתן להימנע מבחירה של נביאים עם אחוזי שגיאה גבוהים יותר.
2. הנביא האופטימלי נבחר 25 פעמים. ניתן לראות כי הגדלת ה-train set אפשרה לנביא האופטימלי להיבחר רבע מהפעמים, מה שמראה שנביאים עם אחוזי שגיאה גדולים יותר ניבאו נכון באחוזים נמוכים יותר וה-true risk שלהם השתקף טוב יותר, כפי שהוסבר ב-1.
3. בכל 100 הניסויים שערכנו בחרנו בנביא שהיה גרוע בפחות מאחוז מהנביא האופטימלי. ההסבר לכך הוא שכאשר יש מספר נביאים שה-true risk שלהם קרובים מאוד אחד לשני, קשה לאלגוריתם להבדיל ביניהם. בפרט, מתוך 500 הנביאים הרנדומליים ישנם נביאים שהגרלנו שאחוזי השגיאות שלהם קרובות עד כדי 0.01 לאחוז השגיאה הנמוך ביותר. כאשר אלו חזו את תוצאות אלף המשחקים, אחוזי השגיאות שלהם נמוכים מאוד ואף עשויים להיות טובים יותר מאחוזי השגיאה של הנביא האופטימלי ולכן האלגוריתם בחר בראשון מבניהם כפי שהוגדר בהוראות.
4. חשבו שגיאות:
   a. Mean Approximation – טעות הקירוב הממוצעת היא בערך 0.003.
   b. Mean Estimation - טעות ההערכה הממוצעת היא בערך 0.0009, לעומת 0.088 במקרה הקודם. בדומה להסבר בסעיף 1, שגיאה זו זניחה ומראה שהאלגוריתם בחר בממוצע את הנביא האופטימלי או נביאים כמעט זהים ב-true risk שלהם לנביא האופטימלי.
5. איך ה-Generalization Error של הנביא שנבחר משתנה בחישוב על-פני ה-Train Set וה-Test Set? מדוע? שגיאת ההכללה קטנה בחישוב על פני ה-Test set לעומת החישוב על פני ה-Train set. זאת כיוון שבדרך כלל ה-empirical risk של הנביא הנבחר יהיה קטן יותר מה-true risk שלו, כלומר סביר שנבחר נביא שניבא בצורה טובה מתוך מקריות. לעומת זאת, כאשר הנביא הנבחר נבדק על פני ה-Test set וצריך להתמודד עם נתונים חדשים, ה-empirical risk יתקרב ל-true risk, כי בדרך כלל כמות הנתונים קטנה יותר והסבירות ששוב יצליח לנבא באותה רמה של הצלחה היא קטנה. לכן, נצפה שעל פני ה-test set יקטן ה-generalization error שמוגדר כהפרש מוחלט בין ה-true risk של נביא מסוים לבין ה-empirical risk של אותו נביא. בדוגמה שלנו כפי שניתן לראות בתוצאות טעות ההכללה לא גדלה על ה-test set לעומת ה-train set, אך ניתן לייחס זאת למקריות או לכך שהיחס בין גודל ה-train ל-test הביא לכך שהנביא הנבחר היה בדרך כלל קרוב מאוד לאופטימלי.

<u>מקרה 5</u>

מחלקת היפותזות טובה, נבחר רנדומלית K in {2, 5, 10, 50} נביאים עם שגיאות שמתפלגות אחיד על התחום [0, 0.2]. נבחר ביניהם בשימוש ב-ERM על פני Train Set של M in {1, 10, 50, 1000} משחקים. נחזור על הניסוי 100 פעמים.

<u>תוצאות ותשובות</u>

1. ERM Approximation Errors (Test Errors) for each (K, M) values

ERM approximation errors

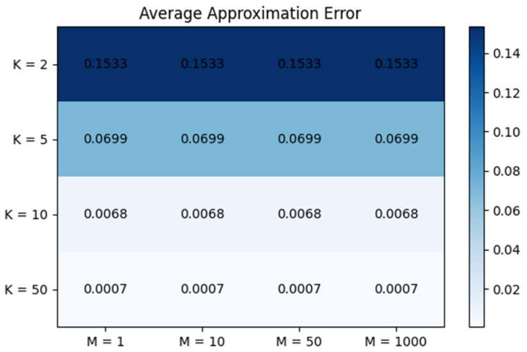
	M = 1	M = 10	M = 50	M = 1000
K = 2	0.1574	0.1593	0.1580	0.1557
K = 5	0.1049	0.0952	0.0823	0.0706
K = 10	0.0233	0.0289	0.0204	0.0068
K = 50	0.1506	0.0927	0.0116	0.0008

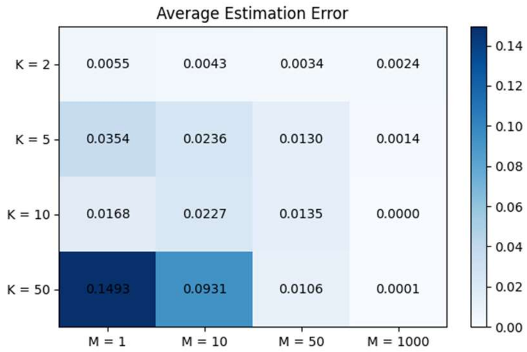
2. Approximation and Estimation Errors for each (K, M) values

Approximation and Estimation errors

	M = 1	M = 10	M = 50	M = 1000
K = 2	(0.1533, 0.0055)	(0.1533, 0.0043)	(0.1533, 0.0034)	(0.1533, 0.0024)
K = 5	(0.0699, 0.0354)	(0.0699, 0.0236)	(0.0699, 0.0130)	(0.0699, 0.0014)
K = 10	(0.0068, 0.0168)	(0.0068, 0.0227)	(0.0068, 0.0135)	(0.0068, 0.0000)
K = 50	(0.0007, 0.1493)	(0.0007, 0.0931)	(0.0007, 0.0106)	(0.0007, 0.0001)

3. הסבירו את התבנית שניתן לראות בטבלה והשוו אותם לתצפיות במקרים 1-4





ניתן לראות שככל שמספר הנביאים K קטן יותר ה-Approximation Error גדול יותר. ההסבר לכך הוא שככל שמספר הנביאים (K) קטן יותר אז יש פחות הזדמנויות לבחור רנדומלית נביא שה-true risk שלו קטן, ואם היה מספר נביאים גדול יותר אז היינו מצפים שגם טווח הערכים (הערכים הקיצוניים) של ה-true risk יגדל ולכן גם יהיה נביא עם true risk קטן יותר. בנוסף, ניתן לראות שככל שמספר המשחקים M גדול יותר, ה-Estimation קטן יותר. זאת כיוון שהסיכוי שלנו לבחור נביא עם true risk נמוך יותר בהתחשב בתוצאות של מספר גדול יותר של ניסויים גדל, כי מחוק המספרים הגדולים, השגיאה הממוצעת על פני יותר משחקים תתקרב לאחוז השגיאה האמיתי, וכך יהיה ניתן להימנע מבחירה של נביאים עם אחוזי שגיאה גבוהים יותר. כמו כן, ניתן לראות בתוצאות טבלת ה-estimation שכאשר מספר המשחקים קטן אז ככל שעולים במספר הנביאים כך ה-estimation גדל, וכאשר מספר המשחקים גדול, ככל שעולים במספר הנביאים כך ה-estimation קטן. ניתן לייחס זאת ליחס מספר המשחקים-מספר הנביאים שמשפיע על טיב ה-estimation במודל.

במקרים 1-4 אנחנו רואים מגמה דומה, ככל שהגדלנו את מספר המשחקים הestimation error קטן. במקרה 1 שבו הערכנו על פני משחק אחד ה-estimation error היה בערך 0.082. במקרה 2 הערכנו את אותם הנביאים על פני עשרה משחקים וקיבלנו Estimation error של בערך 0.036. בנוסף לכך, במקרה 2 הערכנו 500 נביאים רנדומליים (אחוזי שגיאה מתפלגים אחיד על [0,1]) על פני עשרה משחקים וקיבלנו Estimation error של בערך 0.0004 וכשהערכנו 500 נביאים רנדומליים על פני אלף משחקים, קיבלנו Estimation error של בערך 0.0002.

<u>מקרה 6</u>

שתי מחלקות היפותזות מתחרות, כל אחת תייצג שקלול-תמורות Bias-Complexity Tradeoff שונה. מחלקת ההיפותזות הראשונה כוללת 5 נביאים רנדומליים מהתחום [0.3,0.6], מחלקת ההיפותזות השנייה כוללת 500 נביאים רנדומליים מהתחום [0.25,0.6]. גודל ה-Training Set הוא 10 משחקים, גודל ה-validation set וגודל ה-test set הם 1000 משחקים כל אחת. נבחר את הנביא הכי טוב מכל מחלקה באמצעות ERM על ה-Training set. נחזור על הניסוי הזה 100 פעמים.

```
Scenario 6 Results:
5 prophets [0.3,0.6] error rates, 10 games
Average test error of selected prophet: 0.47827999999999987
Number of times best prophet selected: 26
Number of times prophet selected within epsilon: 65
Average approximation error: 0.4488301947294778
Average estimation error: 0.028215032455950764
Average train generalization error: 0.16148987643100565
Average test generalization error: 0.16179999999999997

500 prophets [0.25,0.6] error rates, 10 games
Average test error of selected prophet: 0.3228000000000001
Number of times best prophet selected: 0
Number of times prophet selected within epsilon: 2
Average approximation error: 0.2502055950936596
Average estimation error: 0.07651063426169159
Average train generalization error: 0.3267162293553511
Average test generalization error: 0.3228000000000001
```

<u>תשובות</u>

1. חישוב שגיאות

<u>מחלקת נביאים 1</u>

a. Mean Approximation Error - טעות הקירוב הממוצעת היא בערך 0.448.
b. Mean Estimation Error - טעות ההערכה הממוצעת היא בערך 0.028.

<u>מחלקת נביאים 2</u>

a. Mean Approximation Error - טעות הקירוב הממוצעת היא בערך 0.25.
b. Mean Estimation Error - טעות ההערכה הממוצעת היא בערך 0.076.

2. הסבירו את ה-bias-complexity tradeoffs השונים שכל אחת מהמחלקות מייצגת

במחלקת הנביאים הראשונה שגיאת ה-estimation קטנה יותר (0.028) מאשר במחלקת הנביאים השנייה (0.076), מכיוון שגודלה של המחלקה השנייה מגדיל את הסיכוי לבחירה באמצעות ERM של נביא שה-true risk שלו גבוה יותר אך באופן מקרי הצליח לנבא נכון באחוזים גבוהים. לעומת זאת, במחלקת הנביאים הראשונה יש bias גדול יותר (0.448) לעומת ה-bias במחלקת הנביאים השנייה (0.25), כלומר ה-true risk של הנביא האופטימלי גבוה יותר מאשר במחלקה השנייה.
זאת כיוון שבהגרלת מספר נביאים גדול יותר יש סיכוי גדול יותר להגריל נביא בעל true risk נמוך. הסברים מפורטים יותר בסעיפים הקודמים.