4 בפברואר 2024

1 תרחישים

1.1 תרחיש 1

1.1.1 רקע

בתרחיש זה התבקשנו לייצר שני נביאים ־ אחד עם True Risk של 0.2 והשני עם 0.4.
על שני הנביאים הללו עשינו ניסוי:

1. למידה על סט משחקים בגודל 1 (כלומר משחק יחיד).
2. בהתאם ללמידה, נבחר ב־ERM את הנביא האופטימלי ־ האחד עם השגיאה הקטנה ביותר.
במקרה זה הלמידה תייצר תשובה של 1 או 0, זאת אומרת שהנביא יהיה אבסולוטית צודק או אבסולוטית טועה, אין התפלגות בתוך הסט (אלא רק בין כמה סטים ע"ג כמה ניסויים).
3. לאחר שבחרנו את הנביא האופטימלי בעיני ה־ERM, נעשה עליו בדיקה באמצעות סט בדיקה.
התוצאה של הבדיקה הזו תראה לנו מה השגיאה של הנביא, בעצם יענה לנו על השאלה "כמה מוצלח היה שלב הלמידה", ככה ששגיאה שמתקרבת יותר ל־True Risk של הנביא המוצלח תהיה אופטימלית.

את הניסוי הזה הרצנו 100 פעמים ובדקנו מספר פרמטרים:

1. **שגיאה ממוצעת** - השגיאה הממוצעת על סט הבדיקה בהינתן 100 הניסויים.
2. **בחירות הנביא המוצלח** - כמה פעמים בחרנו את הנביא המוצלח מבין השניים (כלומר האחד עם שגיאה אמיתית של 0.2).
3. **שגיאת הקירוב הממוצעת (Mean Approximation Error)** - השגיאה האמיתית של הנביא המוצלח במהלך כל הניסויים.
4. **שגיאת ההערכה הממוצעת (Mean Estimation Error)** - ההפרש בין הנביא המוצלח לנביא שנבחר, באופן ממוצע על כלל הניסויים.

1.1.2 תוצאות

תוצאות בין 0 ל־1 יוכנסו בדיוק של 3 ספרות אחרי הנקודה.

1. **שגיאה ממוצעת** - 0.224

2. **בחירות הנביא המוצלח** - 87 מ־100 פעמים.

3. **שגיאת קירוב ממוצעת** - 0.2 (קבוע ע"י ההפרש של השגיאה האמיתית של שני הנביאים).

4. **שגיאת הערכה ממוצעת** - 0.026

1.1.3 ניתוח

נשים לב לקורלציה בין תוצאה 1 ל־2, משום שאחוז הבחירה של הנביא המוצלח הוא גבוה, אז גם השגיאה הממוצעת מאוד מתקרבת לסיכון האמיתי של הנביא המוצלח (עד כדי ~0.024).
כמו כן זה מגיע בקורלציה עם תוצאה 4 שהיא מראה לנו מדוע מתקיימת הסטייה בשגיאה הממוצעת מהסיכון האמיתי של הנביא המוצלח.
נשים לב לנקודה חשובה ־ בניסוי שלנו, וכפי שנאמר במסמך התרגיל, עלינו לבחור את הנביא **הראשון** במידה ושני הנביאים צודקים. כלומר, הסדר בו אנחנו בודקים את הנביאים בניסוי חשוב מאוד.
בניסוי הזה בדקנו קודם כל את הנביא המוצלח מבין השניים (עם סיכון 0.2) ורק לאחר מכן את השני (עם סיכון 0.4).
זאת אומרת שיכול להיות שיתכן שקיבלנו תוצאה זהה בשניהם, ובגלל שהסיכון של הנביא המוצלח הוא 0.2 לעומת 0.4 של השני, אז עדיין הדיוק שלנו טוב ולכן קיבלנו שגיאה ממוצעת שלא עולה בהרבה על הסיכון האמיתי של הנביא המוצלח.
אז ננסה לבדוק מה קורה אם מחליפים את הסדר בין שני הנביאים, וקיבלנו ש־

1. **שגיאה ממוצעת** - 0.329

2. **בחירות הנביא המוצלח** - 35 מ־100 פעמים.

3. **שגיאת הערכה ממוצעת** - 0.129

ואכן התוצאות שונות באופן דרמטי, וזה הגיוני מאוד.
הדגימה שלנו בשלב הלמידה היא מאוד (אבל מאוד) מצומצמת, ובגלל שנקבל תוצאות בינאריות 1 או 0, יתכן בסיכוי גבוה ששלב הניסוי לא יעבור באופן חלק (= הניסוי יצא עם שגיאה גבוהה מאוד).
ואנחנו רואים את זה באופן בהיר בתוצאות החדשות הנ"ל, השגיאה הממוצעת מאוד מתקרבת לנביא הפחות מוצלח עם סיכון אמיתי של 0.4.
הניסוי הזה מעלה את השאלה, כפי שלמדנו בהרצאה ־ "מה יקרה אם נגדיל את סט הלמידה?", ועל שאלה זו בדיוק עונים בתרחיש 2.

1.2 תרחיש 2

1.2.1 רקע

התרחיש הזה זהה לתרחיש 1 שקדם לו, מלבד העובדה שנשנה את גודל סט המשחקים ללמידה בכל ניסוי מ־1 (בתרחיש 1) ל־10. נבדוק בתרחיש זה את אותם הפרמטרים כמו בתרחיש 1.
כפי שלמדנו בהרצאה, אנחנו מצפים לראות דיוק טוב יותר בבחירת הנביא המוצלח והתכנסות דרמטית יותר של השגיאה הממוצעת לסיכון האמיתי.

1.2.2 תוצאות

תוצאות בין 0 ל־1 יוכנסו בדיוק של 3 ספרות אחרי הנקודה.

1. **שגיאה ממוצעת** - 0.215
2. **בחירות הנביא המוצלח** - 92 מ־100 פעמים.
3. **שגיאת קירוב ממוצעת** - 0.2 (קבוע ע"י ההפרש של השגיאה האמיתית של שני הנביאים).
4. **שגיאת הערכה ממוצעת** - 0.016

1.2.3 ניתוח

כפי שההשערה שלנו אמרה ־ השגיאה הממוצעת התכנסה עוד יותר ל־0.2 (שגיאת הקירוב). תוצאות הניסוי הזה מתקשרות לנו לעיקרון ML מספר 6 שמופיע בסיכום הקורס ־ שכן באמצעות למידה על מספר פריטים גדול יותר נוכל לקבל תוצאת שגיאה מדויקת יותר עבור השימוש ב־ERM ומכאן שנבחר יותר פעמים את הנביא המוצלח, ולפיכך התוצאות על סט הבדיקה יצאו טובות יותר ויקטינו את שגיאת ההערכה, שניכר שירדה.
כמו בתרחיש 1, נבדוק מה קורה אם מחליפים את סדר הנביאים:

1. **שגיאה ממוצעת** - 0.247
2. **בחירות הנביא המוצלח** - 76 מ־100 פעמים.
3. **שגיאת הערכה ממוצעת** - 0.048

ואכן התוצאות האלו מאששות גם כאן את העובדה שככל שמספר הפריטים עליו אנו לומדים גדול יותר, ככה השגיאות של שני הנביאים בעבור סט הלמידה יצאו קרובות יותר לשגיאה האמיתית שלהם ולפיכך יעזרו ל־ERM לבחור נכונה.

1.3 תרחיש 3

1.3.1 רקע

בתרחיש זה התבקשנו לייצר 500 נביאים, כך שכל אחד בעל סיכון אמיתי שונה בין 0 ל־1. על כל הנביאים ביצענו ניסוי:

1. למידה על סט משחקים בגודל 10.
2. בהתאם ללמידה, נבחר ב־ERM את הנביא האופטימלי ־ האחד עם השגיאה הקטנה ביותר.
3. לאחר שבחרנו את הנביא האופטימלי בעיני ה־ERM, נעשה עליו בדיקה באמצעות סט בדיקה.

את הניסוי הזה הרצנו 100 פעמים ובדקנו מספר פרמטרים:

1. **שגיאה ממוצעת** - השגיאה הממוצעת על סט הבדיקה בהינתן 100 הניסויים.
2. **בחירות הנביא המוצלח** - כמה פעמים בחרנו את הנביא המוצלח מבין השניים (כלומר האחד עם שגיאה אמיתית של 0.2).
3. **שגיאת הקירוב הממוצעת (Mean Approximation Error)** - השגיאה האמיתית של הנביא המוצלח במהלך כל הניסויים.
4. **שגיאת ההערכה הממוצעת (Mean Estimation Error)** - ההפרש בין הנביא המוצלח לנביא שנבחר, באופן ממוצע על כלל הניסויים.
5. **בחירות נביא שחרגו מסף 1%** - בהינתן הסיכון האמיתי של הנביא המוצלח, נספור את כמות הפעמים שבחרנו בנביא עם סיכון גבוה יותר מ־1% מהסיכון האמיתי של הנביא המוצלח.

1.3.2 תוצאות

תוצאות בין 0 ל־1 יוכנסו בדיוק של 3 ספרות אחרי הנקודה.

1. **שגיאה ממוצעת** - 0.151
2. **בחירות הנביא המוצלח** - 0 מ־100 פעמים.
3. **שגיאת קירוב ממוצעת** - 0.009
4. **שגיאת הערכה ממוצעת** - 0.142
5. **בחירות נביא שחרגו מסף 1%** - 97 מ־100 פעמים.

1.3.3 ניתוח

התבקשנו לייצר מחלקת היפותזה גדולה ־ 500 נביאים, כאשר הסיכון האמיתי שלהם נע בין 0 ל־1. כלומר, יש באוסף הנביאים פוטנציאל להמון נביאים גרועים עם סיכון אמיתי גבוה, מה גם שיכולים להיות גם הרבה נביאים שקרובים לסיכון האמיתי של הנביא המוצלח ביותר. מכאן שהתוצאות שנקבל בשלב הלמידה של כל ניסוי יכולות לצאת באופן מקרי טובות יותר או טובות במידה שווה לנביא המוצלח ביותר, ומשום שנבחר את הראשון מבין הטובים, הסיכויים שנבחר את הנביא הטוב ביותר נמוכים מאוד.
הדברים הללו מתחברים לעקרונות ML מספר 7 (עבור הרבה נביאים קרובים לסיכון האמיתי של המוצלח ביותר) ומספר 10 ו־11 (היתרונות והחשיבות של מחלקה קטנה ומוצלחת על פני גדולה ותנודתית).

1.3.4 מקרה נוסף

התבקשנו לייצר מקרה נוסף לתרחיש בו נעבוד עם נביאים בעלי סיכון אמיתי בין 0 ל־0.5. במקרה הזה אנחנו מתגברים חלקית על אחד מהפערים של המקרה המקורי וכפי שתיארנו בניתוח שלו ־ מחלקת היפותזה גדולה עם נביאים גרועים יכולה לייצר שגיאה ממוצעת גבוהה. ולכן כאשר נוריד את הסיכון האמיתי של כל הנביאים לטווח מצומצם יותר ונמוך יותר, כך השגיאות שלנו יהיו טובות יותר (= נמוכות יותר). נבדוק את זה אל מול התוצאות:

1. **שגיאה ממוצעת** - 0.126
2. **בחירות הנביא המוצלח** - 0 מ־100 פעמים.
3. **שגיאת קירוב ממוצעת** - 0.004
4. **שגיאת הערכה ממוצעת** - 0.121

ואכן קיבלנו שגיאות נמוכות יותר, אם כי עדיין לא בחרנו בנביא המוצלח ביותר כי מחלקת ההיפותזה שלנו ככל הנראה עדיין "סובלת" מהרבה נביאים עם סיכון אמיתי נמוך.

1.4 תרחיש 4

1.4.1 רקע

נחזור על תרחיש 3 אך הפעם נתאמן על סט עם 1000 משחקים, לעומת 10 משחקים כמו בתרחיש 3.
בתרחיש זה אנחנו מצפים לראות דיוק גבוה יותר של השגיאות (כלומר התכנסות מדויקת יותר של השגיאה הממוצעת לשגיאת הקירוב הממוצעת, וכן שגיאת הערכה נמוכה יותר).

1.4.2 תוצאות

תוצאות בין 0 ל־1 יוכנסו בדיוק של עד 5 ספרות אחרי הנקודה.

1. **שגיאה ממוצעת** - 0.00135
2. **בחירות הנביא המוצלח** - 82 מ־100 פעמים.
3. **שגיאת קירוב ממוצעת** - 0.00103
4. **שגיאת הערכה ממוצעת** - 0.0004
5. **בחירות נביא שחרגו מסף** 1% - 0 מ־100 פעמים.

1.4.3 ניתוח

כפי שאמרה ההשערה, השגיאה הממוצעת מתכנסת באופן מדויק יותר לשגיאת הקירוב הממוצעת (הסיכון האמיתי של הנביא המוצלח) לעומת תרחיש 3.
הסיבה לכך נובעת ישירות מהעיקרון שלמדנו בהרצאה שככל שנגדיל את סט הלמידה, כך ה־ERM יהיה מדויק יותר (מה שיתן יותר בחירות של הנביא המוצלח ושגיאת הערכה נמוכה יותר), וכך הבדיקה מול סט הבדיקה תהיה מדויקת יותר (מה שיתן שגיאה ממוצעת נמוכה).
עבור שגיאת ההכללה (Generalization Error) של הנביא הנבחר על סט הלמידה אל מול סט הבדיקה, נבין שבמקרה של תרחיש זה השגיאה תהיה מינימלית ביותר וכמעט מתכנסת ל־0.
הסיבה לכך היא שסט הלמידה וגם סט הבדיקה שלנו מורכבים מ־1000 דגימות של משחקים, ולכן ההתפלגות תהיה דומה עד כדי זהה.
זה לא היה המקרה אם סט הלמידה שלנו היה קטן פי כמה מונים מ־1000, כמו בתרחיש 3, ומה שאכן גרם לכך שהשגיאה הממוצעת בתרחיש 3 תהיה גבוהה יותר פי כמה מונים.

1.5 תרחיש 5

1.5.1 רקע

בתרחיש זה התבקשנו לייצר מספר משתנה של נביאים k כך ש־$k \in \{2,5,10,50\}$, כך שכל אחד בעל סיכון אמיתי שונה בין 0 ל־0.2.
על כל k הנביאים ביצענו ניסוי:

1. למידה על סט משחקים בגודל m, כך ש־$m \in \{1,10,50,1000\}$.
2. בהתאם ללמידה, נבחר ב־ERM את הנביא האופטימלי ־ האחד עם השגיאה הקטנה ביותר.
3. לאחר שבחרנו את הנביא האופטימלי בעיני ה־ERM, נעשה עליו בדיקה באמצעות סט בדיקה.

את הניסוי הזה הרצנו 100 פעמים עבור כל m, ובדקנו מספר פרמטרים:

1. **שגיאת הקירוב הממוצעת (Mean Approximation Error)** - השגיאה האמיתית של הנביא המוצלח במהלך כל הניסויים.
2. **שגיאת ההערכה הממוצעת (Mean Estimation Error)** - ההפרש בין הנביא המוצלח לנביא שנבחר, באופן ממוצע על כלל הניסויים.

כלומר, נבצע 100 ניסויים עבור זוגות (k,m) כך ש־$(k,m) \in \{2,5,10,50\} \times \{1,10,50,1000\}$.

1.5.2 תוצאות

בטבלה להלן נציג את התוצאות, כאשר תוצאות בין 0 ל־1 יוכנסו בדיוק של עד 3 ספרות אחרי הנקודה.
הטבלה בנויה כך שבשורות יהיה מספר הנביאים ובעמודות יהיו מספר המשחקים.
בכל תא A הוא שגיאת הקירוב (Approximation) ו־E הוא שגיאת ההערכה (Estimation).

k/m	1	10	50	1000
2	$A\{0.04\}, E\{0.003\}$	$A\{0.124\}, E\{0.006\}$	$A\{0.025\}, E\{0\}$	$A\{0.078\}, E\{0\}$
5	$A\{0.029\}, E\{0.001\}$	$A\{0.005\}, E\{0.045\}$	$A:\{0.013\}, E\{0.008\}$	$A\{0.46\}, E\{0\}$
10	$A\{0.007\}, E\{0.035\}$	$A\{0.029\}, E\{0.065\}$	$A\{0.005\}, E\{0.006\}$	$A\{0.037\}, E\{0.001\}$
50	$A\{0.014\}, E\{0.002\}$	$A\{0.003\}, E\{0.048\}$	$A\{0.0006\}, E\{0.012\}$	$A\{0.006\}, E\{0\}$

1.5.3 ניתוח

ניתן לראות בטבלה מספר דפוסים מעניינים, שראינו כבר בתרחישים הקודמים:

1. **הגדלת מספר המשחקים ללמידה $\Longleftarrow$ הקטנת שגיאת ההערכה:** ראינו זאת כבר בתרחיש 4 מול 3 ותרחיש 2 מול 1, בהם עשינו את אותו הניסוי אך עם מספר עולה של משחקים ללמידה.
גם בטבלה ניתן לראות שבמעבר בין 50 ל־100 משחקים יש שינוי ניכר ודרמטי בהקטנת ה־Estimation Error, כאשר ב־1000 משחקים, בכל המצבים, נתכנס כמעט לשגיאה אפסית.

ניתן לראות את הדפוס הזה גם במעברים בין 10 ל־50 ו־1 ל־10 משחקים, אך יש לעיתים גם עליה, וזאת משיקולים הסתברותיים וכן תלות בשגיאת הקירוב.
נניח במעבר בין $(5,1)$ ו־$(5,10)$ רואים עליה של שגיאת ההערכה, אך זה ככל הנראה נגרם בעיקר בגלל העובדה ששגיאת הקירוב במקרה של $(5,10)$ משמעותית קטנה משגיאת הקירוב ב־$(5,1)$, ולכן כפי שראינו בתרחישים קודמים (ספציפית תרחיש 3 והאירוע הנוסף שבו) ־ קיומם של נביאים בעלי שגיאות קרובות למדי, אך עדיין נביאים מוצלחים יחסית, יכולים לגרום לשגיאת הערכה גבוהה יותר.

2. **קבוצת נביאים מוצלחת** $\Longleftarrow$ **הקטנת שגיאת ההערכה:** בהשוואה לתרחישים 1 עד 4 (בייחוד 1 ו־3), שגיאות ההערכה בתרחיש זה נמוכות מ־0.1.
ככה שגם עבור האירועים עם מספר נמוך יחסית של מספרים קיבלנו שגיאות הערכה נמוכות.
דבר זה נגרם מכך שקבוצת הנביאים היא מוצלחת, עם שגיאות בטווח של $[0,0.2]$.
אם היינו מגדילים את הטווח אז היינו רואים תרחיש דומה לתרחיש 3 המקורי, בייחוד עבור מספר משחקים קטן ומספר נביאים גדול.
גם כאן ניתן לראות את זה, אך בצורה פחות בולטת, לדוגמא בין $(5,10)$ ו־$(50,10)$ נצפית עליה בשגיאת הערכה, וכן בין $(2,50)$ ו־$(50,50)$.

1.6 תרחיש 6

1.6.1 רקע

תרחיש זה בא לבדוק איך שתי מחלקות היפותזה מגדלים שונים וסיכונים אמיתיים בטווחים משתנים ישפיעו על תוצאות אותו הניסוי.
נייצר בתרחיש 2 מחלקות היפותזה, כלומר שתי קבוצות נביאים -

1. קבוצה 1 בעלת 5 נביאים עם סיכונים אמייתים בטווח $[0.3, 0.6]$.
2. קבוצה 2 בעלת 500 נביאים עם סיכונים אמיתיים בטווח $[0.25, 0.6]$.

עבור כל קבוצה נבצע את הניסוי הבא:

1. למידה על סט משחקים בגודל 10.
2. בהתאם ללמידה, נבחר ב-ERM את הנביא האופטימלי - האחד עם השגיאה הקטנה ביותר.
3. לאחר שבחרנו את הנביא האופטימלי בעיני ה-ERM, נעשה עליו בדיקה באמצעות סט בדיקה.

את הניסוי נבצע 100 פעמים עבור כל קבוצת נביאים.

1.6.2 תוצאות

תוצאות בין 0 ל-1 יוכנסו בדיוק של עד 5 ספרות אחרי הנקודה.
תוצאות קבוצה 1:

1. **שגיאה ממוצעת** - 0.406
2. **שגיאת קירוב ממוצעת** - 0.337
3. **שגיאת הערכה ממוצעת** - 0.067

תוצאות קבוצה 2:

1. **שגיאה ממוצעת** - 0.316
2. **שגיאת קירוב ממוצעת** - 0.25
3. **שגיאת הערכה ממוצעת** - 0.068

1.6.3 ניתוח

עיקרון ה-Bias-Complexity Tradeoff אומר שיש Traeoff על מספר הנביאים, כך שיותר נביאים מגדילים את הסיכוי לבחירת נביא טוב יותר, אך כך גם יגדל ה-Estimation Error ושגיאת ההכללה.
ניתן לראות זאת יחסית בבירור ע"י התוצאות למעלה - בקבוצה 1 יש לנו פחות נביאים וה-Estimation Error קטן יותר. בקבוצה 2 קורה ההפך. וזאת ע"פ העיקרון.
ההפרש בין שתי שגיאות ההערכה עבור שתי הקבוצות קטן יחסית, אך הוא ככל הנראה נגרם ע"י הרנדומיות שבתרגיל, שכן בהרצה נוספת קיבלתי תוצאות שמשקפות טוב יותר את העיקרון:
תוצאות קבוצה 1:

1. **שגיאה ממוצעת** - 0.481
2. **שגיאת קירוב ממוצעת** - 0.447
3. **שגיאת הערכה ממוצעת** - 0.036

תוצאות קבוצה 2:

1. **שגיאה ממוצעת** - 0.315
2. **שגיאת קירוב ממוצעת** - 0.25
3. **שגיאת הערכה ממוצעת** - 0.066