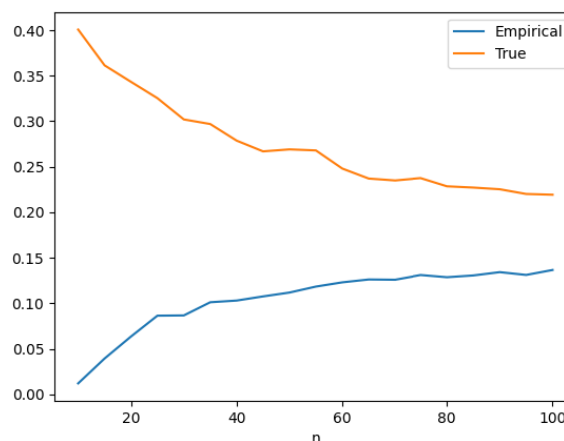


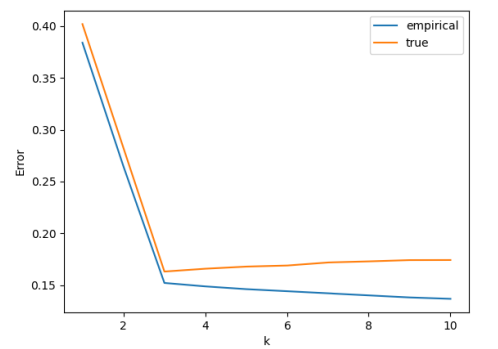
המשך – גרפים

סעיף ב'



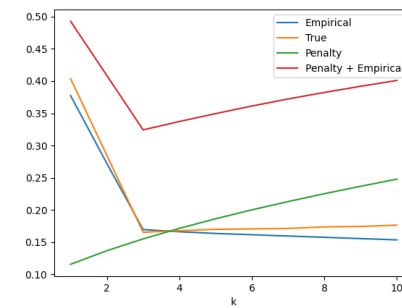
ניתן לראות מהגרף כי השגיאה האמיתית יורדת ככל שגודל המדגם גדל כי ככל שכמות הדוגמאות גדלה אנחנו מתקרבים יותר להסתברות האמיתית של העולם P .
השגיאה האמפירית עולה ככל שגודל המדגם עולה מפני שככל שיש יותר דגימות הסיכוי לקבל תוצאות קיצוניות גדל ולכן השגיאה עולה בהתאם.
ניתן גם לראות שהערכים מתקרבים אחד לשני וזה גם מסתדר כי התוחלת של השגיאה האמפירית היא השגיאה האמיתית ולכן ככל שנגדיל את גודל המדגם הערכים יתקרבו אחד לשני.

סעיף ג'



ניתן לראות מהגרף כי השגיאה האמפירית יורדת ככל ש k גדל, ואילו השגיאה האמיתית מגיעה לתחתית ב $k=3$ ואז מתחילה לעלות. קיבלתי כי $k^* = 10$, אבל נשים לב כפי שצינתי כבר עבור $k=3$ מקבלים את השגיאה האמיתית הנמוכה ביותר, ולכן הבחירה של k^* לא מוצלחת. מצב זה נוצר כתוצאה מאימון יתר של הנתונים שלנו שגורם לרגישות יתר לרעשים ותוצאות קיצוניות.

סעיף ד'



נשים לב שהקנס עולה ככל ש k גדל וזה הגיוני כי המימד גדל כלומר ההיפותזה מסובכת יותר ולכן נצטרך לשלם קנס גבוה יותר, כמובן שניתן גם לראות את זה מחישוב הקנס שיש יחס ישיר בין הקנס ל k .
הערך המינימלי של ה $\text{penalty} + \text{empirical error}$ מתקבל עבור $k=3$ וזהו זהה לתוצאה מהסעיף הקודם.

סעיף ה'

```
[(0.00021516523452441705, 0.19929980245149892), (0.400244140438411, 0.5993662634659548), (0.800076854255245, 0.9977186127349789)]  
best k is 3  
  
Process finished with exit code 0
```

לאחר הרצת האלגוריתם קיבלנו כי ההיפותזה הכי טובה התקבלה עבור $k=3$.
ללא הרצת ה validation היינו אולי יכולים לקבל היפותזה עם שגיאה אמפירית נמוכה יותר כי שראינו בסעיף קודם עם k^* והיינו בוחרים ב k הלא נכון. נשים לב גם שהתוצאה קרובה מאוד להיפותזה האופטימלית שראינו בסעיף א'.