

מודלים לא לינאריים בחקר ביצועים – תרגיל בית 9

31 במאי 2021

הנחיות להגשה

- הגשת התרגיל היא עד יום ראשון, 13.6.21, בשעה 12:00 בצהריים.
- יש להגיש שני קבצים עם שם קובץ HW##_ID1_ID2:
- PDF המכיל את תשובותיכם כולל קודים ופלט.
- ZIP המכיל את כל הקודים.

שאלה 1

בשאלה זו ננתח את התנהגותן של מניות הכלולות במדד האמריקאי S&A 500 בשנת 2016. נתמקד בניתוח מחיר הסגירה של כל מניה על פני שנה, שהוא מחיר המניה בסוף כל יום מסחר.

א. הורידו את הקובץ prices.csv המצורף לתרגיל הבית, בו נמצאים מחירי המניות בפורמט CSV. הריצו את שורות הקוד הבאות, צרפו את הפלט שמתקבל והסבירו מה מבצעות השורה השנייה והשלישית.

```
import pandas as pd
df = pd.read_csv('prices.csv')
df.head(5)
```

ב. הריצו את קטע הקוד המצורף, צרפו את התרשים שמתקבל, והסבירו מה מבצעת כל אחת משורות הקוד.

```
mask = df['date'].apply(lambda x: x[:4] == '2016')
df = df[mask]
df = df[df['symbol'] == 'AAPL'].reset_index()
apple_close_prices = df.close
apple_close_prices.plot()
plt.show()
```

ג. הורידו את הקובץ securities.csv המורף לתרגיל הבית וכיתבו פונקציה המופעלת באופן הבא

```
symbols, prices, sectors = load_shares()
```

הפונקציה טוענת נתונים מהקבצים prices.csv ו-securities.csv מידע על מניות עליהן קיימים נתונים על כל ימי המסחר בשנת 2016, ומחזירה שלושה משתנים:

- הרכיב ה- i ב-symbols הוא מחרוזת עם שם המנייה ה- i .
- הרכיב ה- i ב-sectors הוא מחרוזת עם שמו של המגזר העסקי של המנייה ה- i . המגזר העסקי של כל מניה מופיע בקובץ securities.csv בעמודה GICSSector.
- prices[i,:] הוא וקטור של מחירי הסגירה של המניה ה- i בכל ימי המסחר בשנת 2016.

ד. כתבו פונקציה המופעלת באופן הבא

```
proj = pca_project(X, k)
```

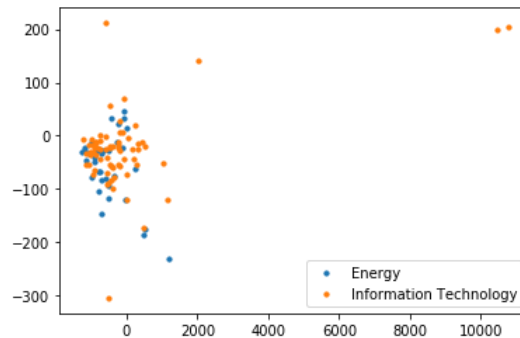
הפונקציה מקבלת וקטורים בשורת המטריצה $X \in \mathbb{R}^{n \times m}$ ומטילה אותה על k וקטורי הבסיס הראשוניים המחושבים ע"י אלגוריתם PCA. הפונקציה מחזירה מטריצה כך ש-proj[i,j] היא ההטלה של $X[i,:]$ על הציר j . אין להשתמש בסעיף זה בפונקציות מובנות למימוש PCA ויש לממש את האלגוריתם בעצמכם.

ה. כתבו פונקציה המופעלת באופן הבא

```
plot_sectors(proj, sectors, sectors_to_plot)
```

הפונקציה מייצרת תרשים של נקודות במישור שנמצאות בשורות המטריצה proj, כאשר כל נקודה מייצגת מניה. המגזר העסקי של כל מניה נמצא במשתנה sectors והפונקציה יוצרת את התרשים רק עבור המניות במגזרים שנתונים במשתנה sectors_to_plot. לצורך בדיקת הקוד עד כה, קטע הקוד הבא אמור ליצור את התרשים המופיע לאחריו

```
_, prices, sectors = load_shares()
proj = pca_project(prices, 2)
plot_sectors(proj, sectors, ['Energy', 'Information Technology'])
```



- ו. בתרשים הקודם לא ניתן להבחין בהבדל בין התנהגותן של מניות בשני מגזרים אלה. הסיבה לכך היא ההבדל בקנה המידה של כל מניה. שתי חברות בעלות אותו שווי ואף התנהגות דומה יכולות להיראות שונה בתרשים, זאת כי אחת מהן בחרה לחלק את השווי שלה למעט מניות יקרות ואילו השנייה בחרה בהרבה מניות זולות. בעיה זו נקראת Data Scaling. אחת הדרכים להתגבר עליה כאשר מדובר בנתונים פיננסיים היא לא לייצג את המניה באמצעות וקטור המחיר $p \in \mathbb{R}^m$ אלא באמצעות הווקטור

$$(\ln(p_{i+1}) - \ln(p_i))_{i=1}^{m-1}$$

חיזרו על הניסוי מהסעיף הקודם, רק שהפעם הזינו לפונקציה pca_project מטריצה שמורכבת מווקטורים בתצורה החדשה. צרפו את הקוד והתרשים שמתקבל.

ז. אחת הדרכים הנפוצות להורדת סיכונים היא לפזר את ההשקעות במגזרים שונים, כך שמגמות במגזר מסויים ישפיעו במידה מועטה על מגמות במגזרים האחרים. תחת הנחה זו, האם כדאי להשקיע ב־זמנית בפיננסים ובטכנולוגיות מידע? מה לגבי השקעה ב־זמנית באנרגיה, טכנולוגיות מידע ונדל"ן? צרפו תרשימים שמגבים את החלטותיכם.

ח. כיתבו קטע קוד היוצר תרשים של כלל המניות מכלל המגזרים באמצעות הפונקציות אותן כתבתם בשאלה זו. בתרשים שיתקבל תבחינו במניה אחת בולטת. מיצאו את המניה (ה־symbol שלה) וצרפו תרשים של מחיריה לאורך השנה, לעומת מניה "טיפוסית" כלשהי. מה ניתן ללמוד מהתרשים? צרפו את שני התרשימים ואת קטעי הקוד.

שאלה 2

תהי נקודה $\mathbf{x}^0 \in \mathbb{R}^n$ ויהי $r > 0$. נגדיר את הקבוצה $B(\mathbf{x}^0, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}^0\| \leq r\}$. הסבירו מדוע $P_{B(\mathbf{x}^0, r)} : \mathbb{R}^n \rightarrow B(\mathbf{x}^0, r)$ קמורה וסגורה ומיצאו נוסחה לחישוב הפונקציה $P_{B(\mathbf{x}^0, r)}$.