

# ML Course

## FALL SEMESTER 2021

### Exercise 6: Pandas, Scikit-Learn, Optuna Tutorial

---

אנו חוזרים לשאיפותינו מהתרגיל השני, כלומר להרוויח קצת כסף בבורסה. לשמחתינו ישנם בידינו נתונים עבור מספר מניות. בנוסף, מצוי בידינו גם מודל מתמטי המתאר את התנהגותן. לצערנו, תנועת המניות כוללת מרכיב רעש אקראי (הכסף הרי אינו צומח על העצים).

המודל המתאר את התנהגות המניות טריוויאלי יחסית<sup>1</sup>: המחיר של כל אחת מורכב ממקטעים (regimes) של תנועה בראונית (Brownian motion) לסירוגין, כאשר אורכם של המקטעים משתנה ואינו ידוע לנו (אך ציפור קטנה לחשה לנו כי הוא נע בין 20 ל-40 ימים). אנו כן יודעים כי המקטעים משתנים לסירוגין ממצב "עליה" (עם ממוצע רווח יומי של 0.7% וסטיית תקן 1%) למצב "ירידה" (הפסד יומי 0.4% וסטיית תקן 1%).

בנוסף אנו יודעים כי המניות כולן "מחויבות" לעקוב אחר המקטעים הנ"ל (הרי הכלכלה משותפת) אך לכ"א מהן השהייה שונה ביחס לבורסה (למשל, אחת מתחילה את המקטע הראשון 5 ימים לפני חברתה, והפרש זה נשאר קבוע לאורך הסימולציה).

---

<sup>1</sup> כלומר יחסית למציאות.

---

## שלב 1:

אנו מעוניינים לסחור במניית Apple (סימון AAPL). בשלב הראשון, ננסה להפיק מהתנועה ההיסטורית של המניה חוקי החלטה עבור מסחר.

🙏 נחלץ (עבור כל יום מסחר) את ההחזר של 10 ימי המסחר האחרונים של AAPL (כלומר את היחס בין מחיר המניה באותו היום לבין מחירה 10 ימים קודם לכן). המאפיין קרוי בחוברת בשם AAPL\_back\_gain\_10.

🧠 על סמך המאפיין הנ"ל, מה חוק ההחלטה המאפשר לנו לקבוע בהסתברות של לפחות 95% כי המניה עברה ממצב מסוים (למשל "עלייה") למצב הנגדי? בחישוב ניתן להתעלם ממספר גורמים "זניחים" (כלומר לא נוחים לחישוב).

### [הפתרון נמצא בתחילת Section C במחברת]

🙏 נסמלץ את תהליך הקניה + מכירה של המניה, כאשר פקודת קניה / מכירה תינתן כאשר אובחן שינוי פאזה לפי הסעיף הקודם (כלומר בהסתברות של 95% בקירוב).

### [הסימולציה מתרחשת בתחתית Section A]

## שלב 2:

ידוע כי AAPL נמצאת בפיגור פאזה של 5 ימים אחרי "השוק". כעת קיבלנו נתונים עבור מניה נוספת (שמה "AUX\_0"), ובנוסף אנו יודעים כי מניה זו נסחרת ללא פיגור פאזה (ולכן מקדימה ב-5 ימים את AAPL). אנו מעוניינים להשתמש בנתוני AUX\_0 רק כעזרים להחלטות הקניה / מכירה אותן אנו

---

מבצעים על AAPL. למען הפשטות, נתון כי ברגע תחילת הסימולציה כל המניות נמצאות במצב עליה.

🧠 אם עומדת לרשותינו גם היסטוריית ההחזרים של AUX\_0, כיצד ישתנה מנגנון ההחלטה?

👧 נגדיר פונ' מטרה: היחס שבין מחיר המניה הממוצע ב-10 הימים העוקבים (עבור כל יום מסחר בנפרד כמובן) לבין מחירה ביום המסחר הנוכחי. עזרו בבקשה לבתיה לקודד את פונ' המטרה במחברת!

### **[נא למלא את החסר בתחילת Section B במחברת]**

🙏 בדקו את גרף האינטראקציות (כלומר scatter plot של כל הדוגמאות) בין ההחזר העתידי של AAPL לבין המאפיין ההיסטורי שלה (הקוד מצוי בחוברת התרגיל). חזרו על הבדיקה כאשר מחליפים את המאפיין ההיסטורי של AAPL במאפיין המקביל של המניה AUX\_0. מה ניתן לומר ממבט מהיר בשני הגרפים? בדקו את הקורלציות גם באופן נומרי (פונ' corr()).

### **[הגרפים נמצאים ב-Section B במחברת]**

🙏 ניגש לקודד ולסמלץ את מנגנון ההחלטה הנ"ל באמצעות פונ' העזר rollout(). מה ניתן לומר על ביצועי המערכת?

### **[הסימולציה נמצאת ב-Section C]**

---

### שלב 3:

נבצע כעת רגרסיה לינארית על המאפיין של המניה AUX\_0 כפי שהוגדר למעלה, עם פונ' המטרה שהוגדרה וקודדה ע"י בתיה.

🧠 בדקו את ביצועי המערכת כאשר פקודות הקניה והמכירה מוכתבות ע"י פלט הרגרסיה. את הספים עבור קבלת ההחלטות נקבע ידנית ("buy\_thresh" ו-"sell\_thresh" במחברת).

- איך הביצועים בהשוואה לסעיפים הקודמים?
- איך ניתן להסביר את ההבדלים?

**[הגרפים המתארים את התהליך נמצאים ב-Section D במחברת]**

### שלב 4:

מטרתנו בשלב זה: אופטימיזציה של פרמטרי המערכת, כלומר הספים אותם קבענו ידנית בסעיף הקודם. ניעזר בספריית Optuna עבור התהליך. ספרייה זו מספקת פתרון נוח ומהיר עבור אופטימיזציה בהיעדר נגזרות, כלומר black box optimization.

**[את תהליך האופטימיזציה ניתן למצוא ב-Section E במחברת]**

🧒 בשלב האחרון, "נשחק" קצת עם המערכת אותה יצרנו.

נא לבצע את הפעולות המופיעות בסיום המחברת.



בהצלחה!