

הרצאה 1 - מבוא

השם מערכות לומדות יכול מאד להטעות, שכן השיטות שבהם משתמשים כיום רחוקות מאד מהלמידה שאנו מכירים מחיי היום יום.

נדגים כיצד עובדות רוב השיטות בתחום בעזרת דוגמאות.

3, 6, 12, ?

1, 1, 2, 3, 5, 8, ?

5, 10, 7, 12, 9, 14, ?

3, 6, 12, 24

$$\Rightarrow n_i = 3 \times 2^i$$

1, 1, 2, 3, 5, 8, 13

$$\Rightarrow n_i = n_{i-1} + n_{i-2}$$

5, 10, 7, 12, 9, 14, 11

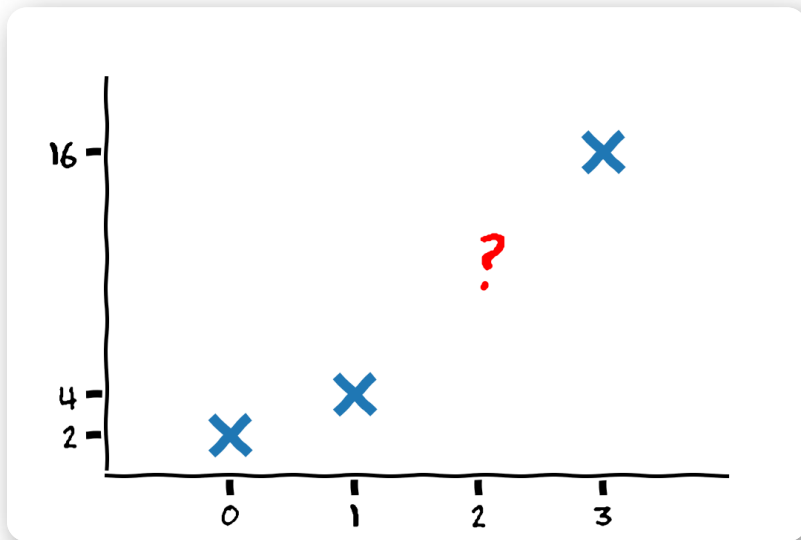
$$\Rightarrow n_i = \begin{cases} n_{i-1} + 5 & i \text{ is odd} \\ n_{i-1} - 3 & i \text{ is even} \end{cases}$$

כיצד אנו פותרים תרגילים כאלה?

- נחפש חוקיות (מודל).
- לרוב נחפש את מודל מתוך אוסף של מודלים מוכרים.
- יתכן יותר ממודל אחד מתאים.
- נעדיף מודל פשוט על פני מודל מסובך (התער של אוקאם).
- מודל יכול להכיל פרמטר שיש לקבוע על פי הדוגמאות (למשל ה-2 וה-3 בדוגמא הראשונה).
- לרוב, ככל שהמודל "מסובך יותר" נצטרך יותר דוגמאות.

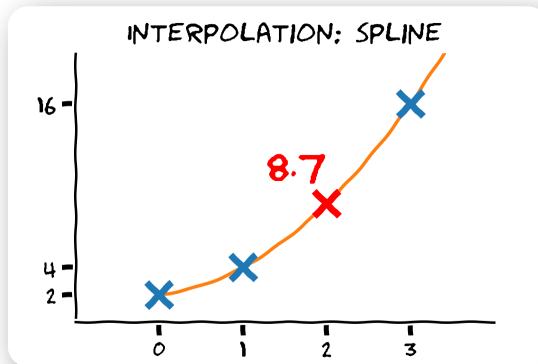
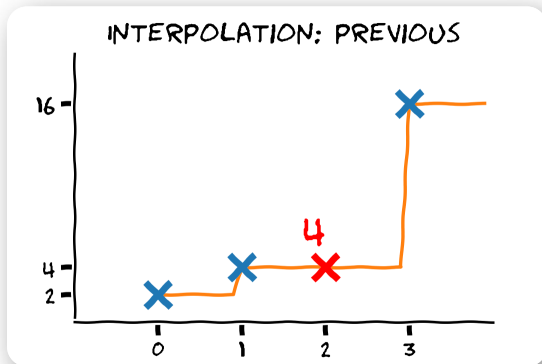
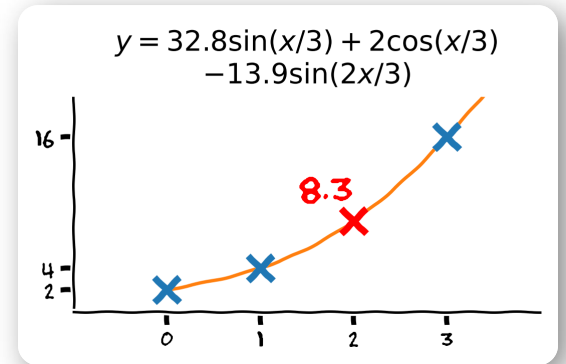
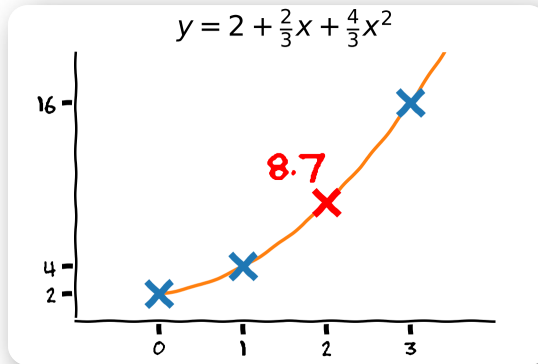
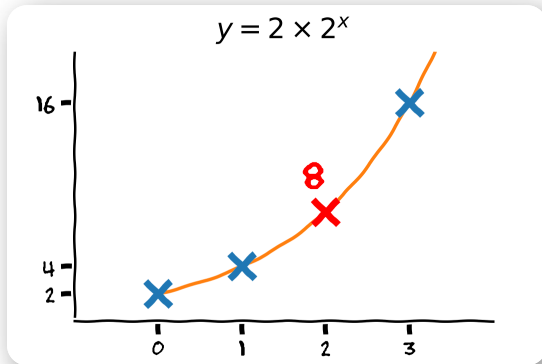
- למצוא מודל שמתאים לדוגמאות זה קל.
- לבחור מבין כל המודלים את המודל הנכון, זה קשה.
(ולרוב בלתי אפשרי)

נסתכל על הדוגמא הבאה



2, 4, ?, 16

מודלים אפשריים



המודל צריך לדעת להכליל

- יש הרבה מודלים שיכולים להתאים לדוגמאות.
- המטרה היא למצוא מודל שידע להכליל למקרים שעוד לא ראינו.
- כדי לבחור את המודל המתאים עלינו להשתמש בידע הקודם שיש לנו.

בעבור המקרה של: 2, 4, ?, 16

מהניסון הקודם שלנו, אנו יודעים שבעיות כאלה נפוץ להשתמש בסדרה הנדסית. לכן המודל הסביר ביותר הוא:

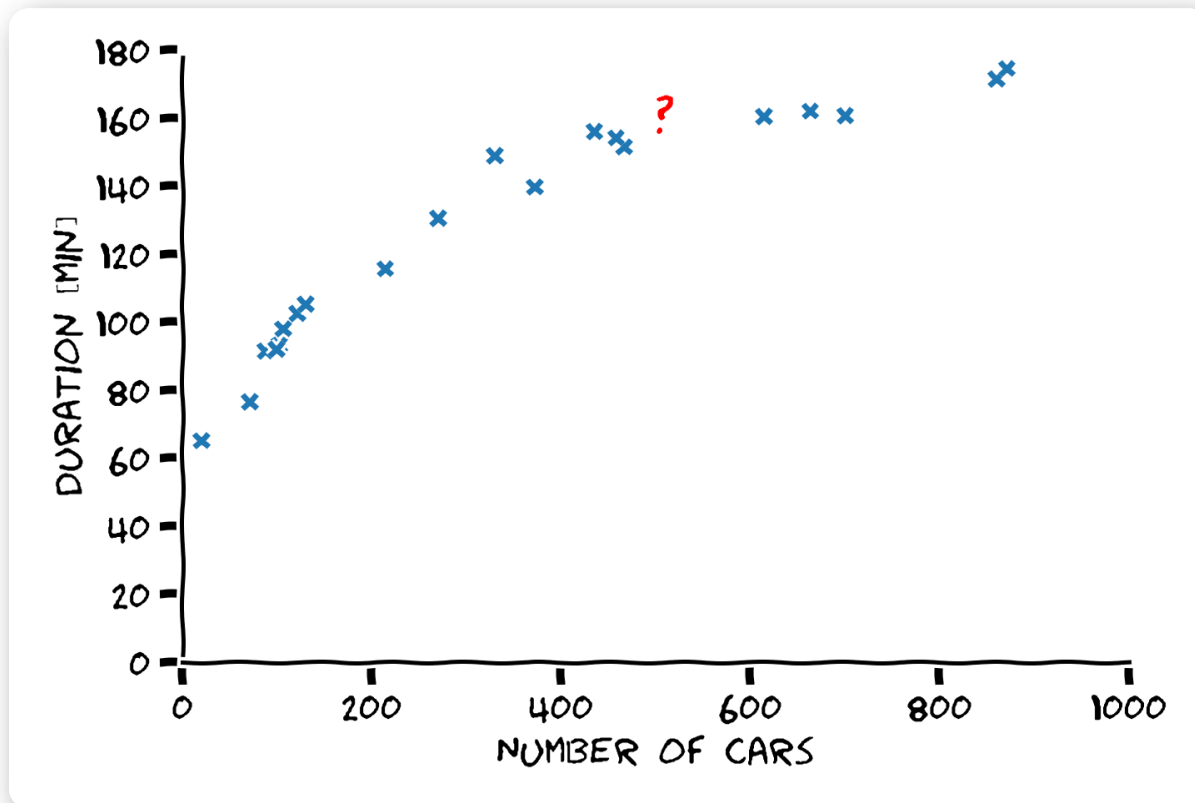
$$n_i = 2 \times 2^i$$

- בדוגמא הצגנו מערכת דטרמיניסטית (לא אקראית). זאת אומרת, שבמקום ה־ i ישב מספר מסויים קבוע.
- בפועל, ברוב במערכות שאותם נרצה למדל יהיה רכיב סטוכסטי (אקראי). זאת אומרת, שיתכן שבעבור פרמטרים זהים נקבל התנהגויות שונות.

מערכות אקראיות - דוגמא

חיזוי זמן נסיעה על פי העומס בכביש (מספר המכוניות על הכביש).

מכיוון שזמן הנסיעה תלוי בעוד הרבה גורמים אחרים חוץ מהעומס, ניתן לקבל זמנים שונים בעבור אותו עומס.



המודל גם יטעה לפעמים

כשהמערכת היא אקראית, או מאד מורכבת, לא נצפה למצוא מודל שתמיד צודק. במקום זאת נחפש מודל שטועה כמה שפחות.

במקרים כאלה נעזר בתורת ההסתברות על מנת לתאר את הבעיה והפתרון.

אז איך עושים את זה?

הרעיון מאחורי רוב השיטות במערכות לומדות הוא זהה:

- 1. נגדיר קריטריון מתמטי אשר מודד עד כמה מודל מסויים מצליח לבצע את המשימה.**
- 2. נבחר משפחה רחבה של מודלים בתקווה שלפחות אחד מהם יהיה מוצלח מספיק.**
- 3. נחפש מבין כל המודלים במשפחה את המודל המוצלח ביותר.**

(הרעיון פשוט, הביצוע קצת פחות).

מרבית הקורס יעסוק בשיטות לביצוע שלושת השלבים האלו.

אנו נבחר לרוב לייצג את המשפחה של המודלים בעזרת **מודל פרמטרי** (פונקציות בעלות מבנה קבוע עד כדי כמה פרמטרים שאותם ניתן לשנות)

דוגמאות

- פונקציות לינאריות: $f_{\theta}(x) = \theta_1 + \theta_2 x$
- כל הפולינומים עד סדר 3: $f_{\theta}(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$
- קומבינציה לינארית של פונקציות: $f_{\theta}(x) = \theta_1 e^x + \theta_2 \sin(x)$
- משהו אחר: $f_{\theta}(x) = \theta_1 e^{-(x-\theta_2)^2/\theta_3}$
- רשת נוירונים.

איך נדאג שהמודל הפרמטרי שלנו יכיל את המודל האופטימאלי?

אנחנו לא!

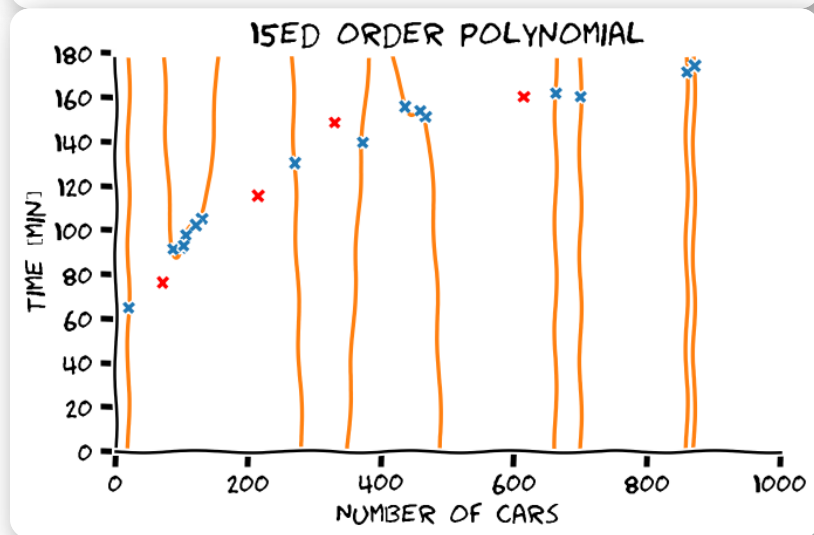
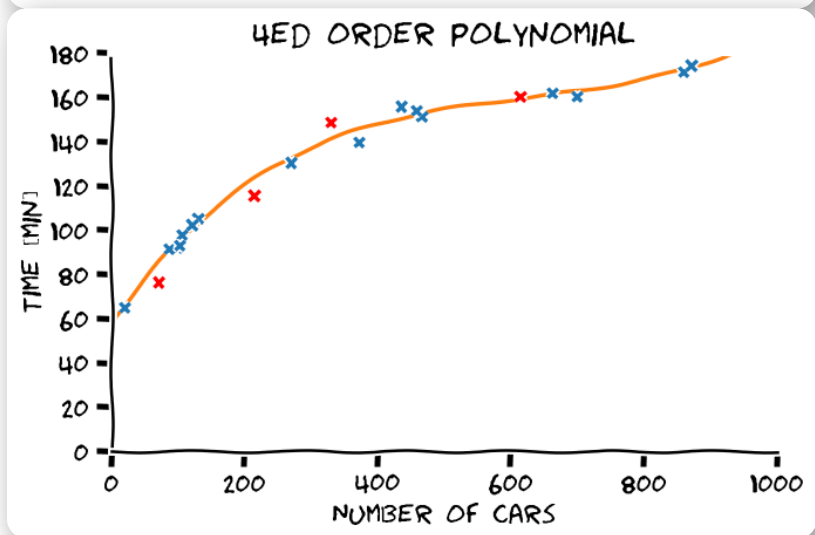
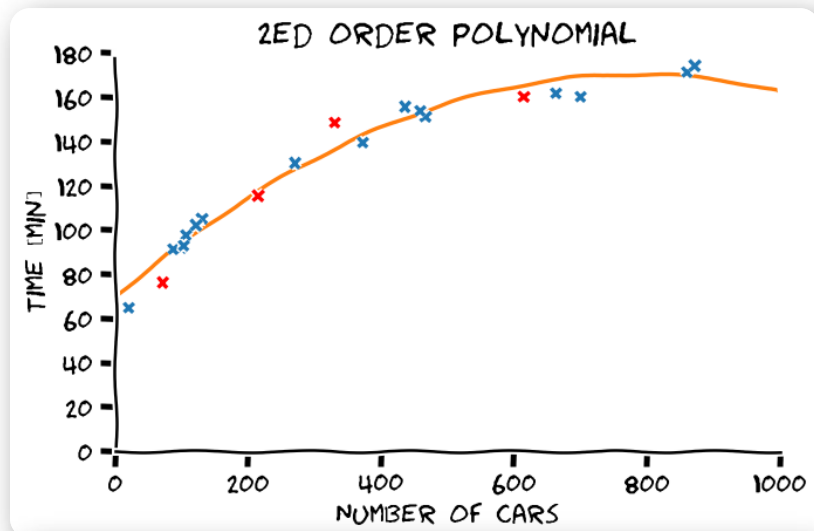
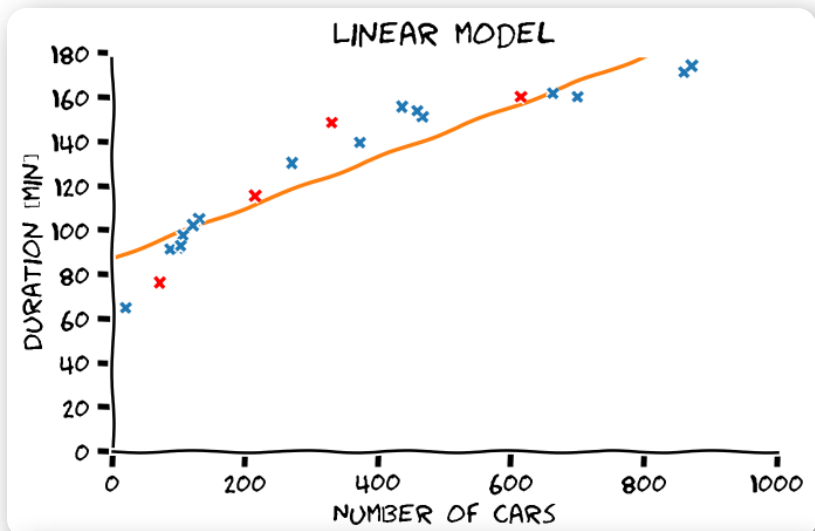
במערכות מורכבות אנחנו כנראה אף פעם לא נוכל למצוא את הפתרון האופטימאלי לבעיה. אנחנו נשאף להגיע כמה שיותר קרוב עליו.

כל המודלים טועים אבל חלק שימושיים

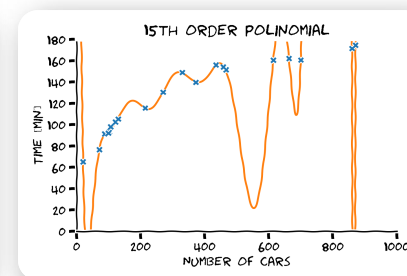
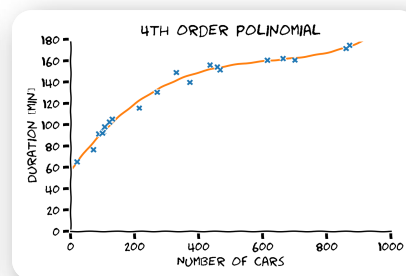
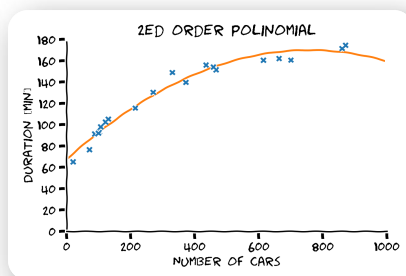
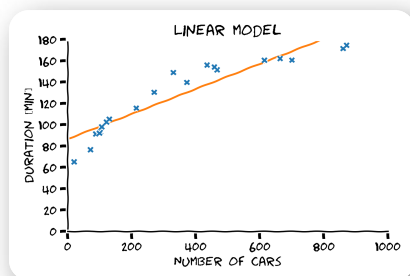
George E. P. Box

בחירת המודל הפרמטרי (משפחת המודלים)

הבחירה של המודל הפרמטרי תשפיע מאד על הפתרון שנקבל



כבר אמרנו קודם, הבעיה היא בעיית הכללה



- סדר פולינום גבוה יותר -> התאמה טובה יותר לדוגמאות.
 - התאמה טובה יותר לדוגמאות לא מעידה על הכללה טובה יותר.
 - את המודל הפרמטרי יש לבחור על סמך ידע קודם.
- מההיכרות שלנו עם הבעיה אנו מצפים שפונקציית המיפוי תהיה מונוטונית עולה וגם שלא תשתנה בפראות.
- מבין האופציות הנ"ל, פולינום מסדר 4 הוא הפשרה הטובה ביותר בין ההתאמה לדוגמאות והידע הקודם.

הרעיון של בניית מודל מתמטי לצורך תיאור של מערכת או לתהליך כל שהוא, הוא למעשה אחד הרעיונות הבסיסיים עליו מושתתים רוב תחומי ההנדסה והמדעים.

- אנו נשתמש בשם **prior knowledge** (או בקיצור **prior**) בכדי להתייחס לידע המוקדם.
- אנו נשתמש בשם **data** בכדי להתייחס לדוגמאות / תצפיות.

- כאשר המערכת פשוטה, ויש לנו הבנה טובה שלה נסתמך בעיקר על הידע הקודם.
- במקרים אחרים נסתמך הרבה יותר על הdata.

לדוגמא

מודל המבוסס על ידע קודם

מכונית יוצאת מחיפה לתל אביב במהירות קבועה ידועה.
(מודל: מיקום = זמן \times מהירות).

מודל המצריך שימוש בdata

מהירות המכונית תלויה במשתנים כגון אופי הנהג ומצב הכביש. ניתן לנסות לבנות מודל על סמך מידע מנסיעות קודמות.

Data vs. Prior

- כאשר המערכת פשוטה, ויש לנו הבנה טובה שלה נסתמך בעיקר על הידע הקודם.
- במקרים אחרים נסתמך הרבה יותר על הdata.

אילוסטרציה

What should the model be based on?

Trying to predict the location of a car...



איך זה מתקשר למערכות לומדות?

הגדרה פורמלית

התחום של מערכות לומדות עוסק באלגוריתמים אשר מנסים להשתמש במידע זמין על מנת לשפר את הביצועים של מכונה במשימה כל שהיא.

בפועל התחום של מערכות לומדות מתעסק בעיקר במקרים בהם אין הרבה ידע מקדים ובניית המודל נעשית בעיקר על סמך הdata.

What should the model be based on?

Trying to predict the location of a car...



The regime of machine learning problems (according to the definition)

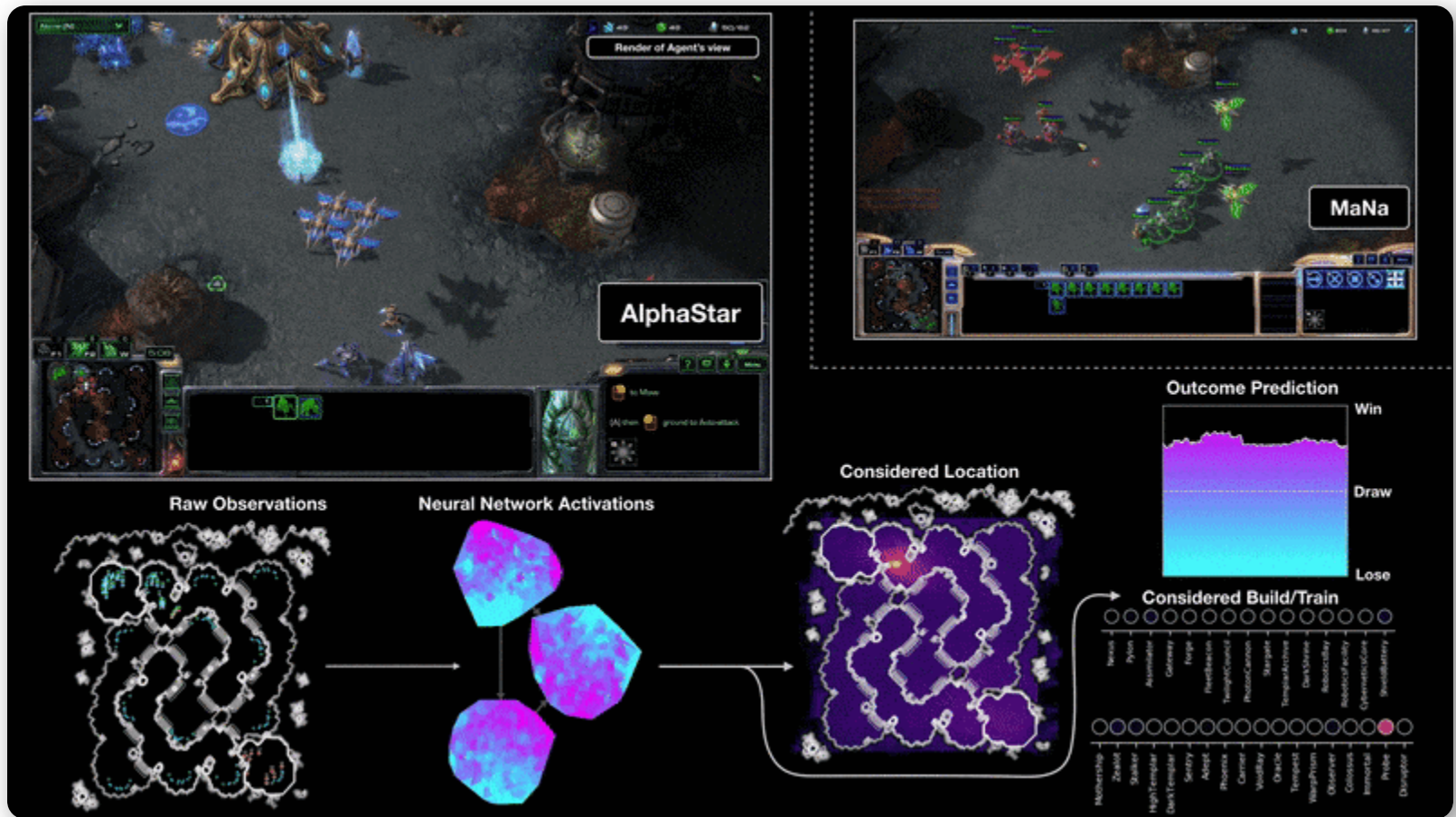
Where typical machine learning problems live

**הנה כמה דוגמאות לדברים שאותם
מערכות לומדות יכולות לעשות:**

לנהוג במכונית מירוץ



לניצח בני אדם במשחקי לוח ומחשב



DeepMind's AlphaStar

לכתוב בלוגים

**Nothing but Words by Liam Porr
(and Open AI's GPT-3)**

לייצר תמונות ריאליסטיות



<https://thispersondoesnotexist.com/>

לנהל שיחות טלפון.



מערכות לומדות בחיי היום יום

- מערכות עזר לנהיגה (mobileye).
- זיהוי הונאות בכרטיסי אשראי.
- סינון דואר.
- שיפור תוצאות חיפוש (גוגל).
- התאמת תוכן למשתמש.
- עוזרות וירטואליות (Siri, Alexa, Cortana, Google Now).

מונחית (Supervised)

יש בידינו דוגמאות של קלט ופלט ממיפוי כל שהוא, ואנו מעוניינים להכליל את הדוגמאות למקרה הכללי.



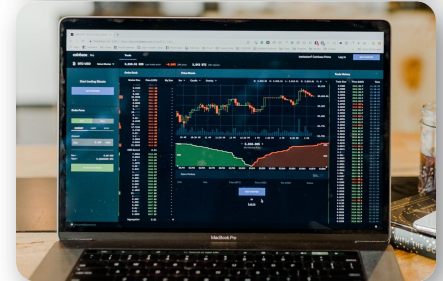
לא מונחית (Unsupervised)

יש בידינו אוסף של דוגמאות כלשהם ואנו מנסים ללמוד את המאפיינים שלהם.



מקוונות (Online)

אנו מעוניינים להמשיך לעדכן את המודל שלנו בעקבות מידע שממשיך להגיע באופן שוטף.



מחיזוקים (Reinforcement)

אנו מאפשרים לאלגוריתם לבצע אינטרקציה עם המערכת וללמוד מהמשוב שהוא מקבל ממנה (ניסוי וטעיה).



סוגי בעיות למידה - דוגמאות

מונחית (Supervised)

זיהוי אובייקטים בתמונה, סינון דואר זבל, סיוע באיבחון רפואי.



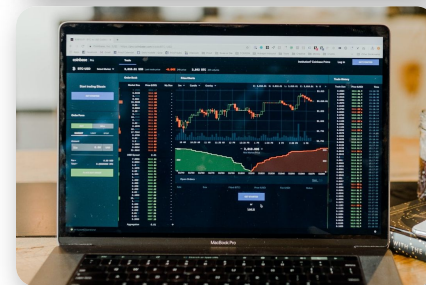
לא מונחית (Unsupervised)

ייצור דוגמאות חדשות על סמך ישנות (תמונות, מוזיקה), שינוי מאפיינים (קול של אדם, פנים), דחיסה.



מקוונות (Online)

סינון דואר זבל עם עידכון על כל דואר חדש שמגיע. חיזוי מחירי מניות עם עדכון על כל מידע חדש שמגיע.



מחיזוקים (Reinforcement)

נהיגה אוטונומית, משחק שח, הליכה. ברוב המקרים האלגוריתם יתאמן תחילה על סימולטור.



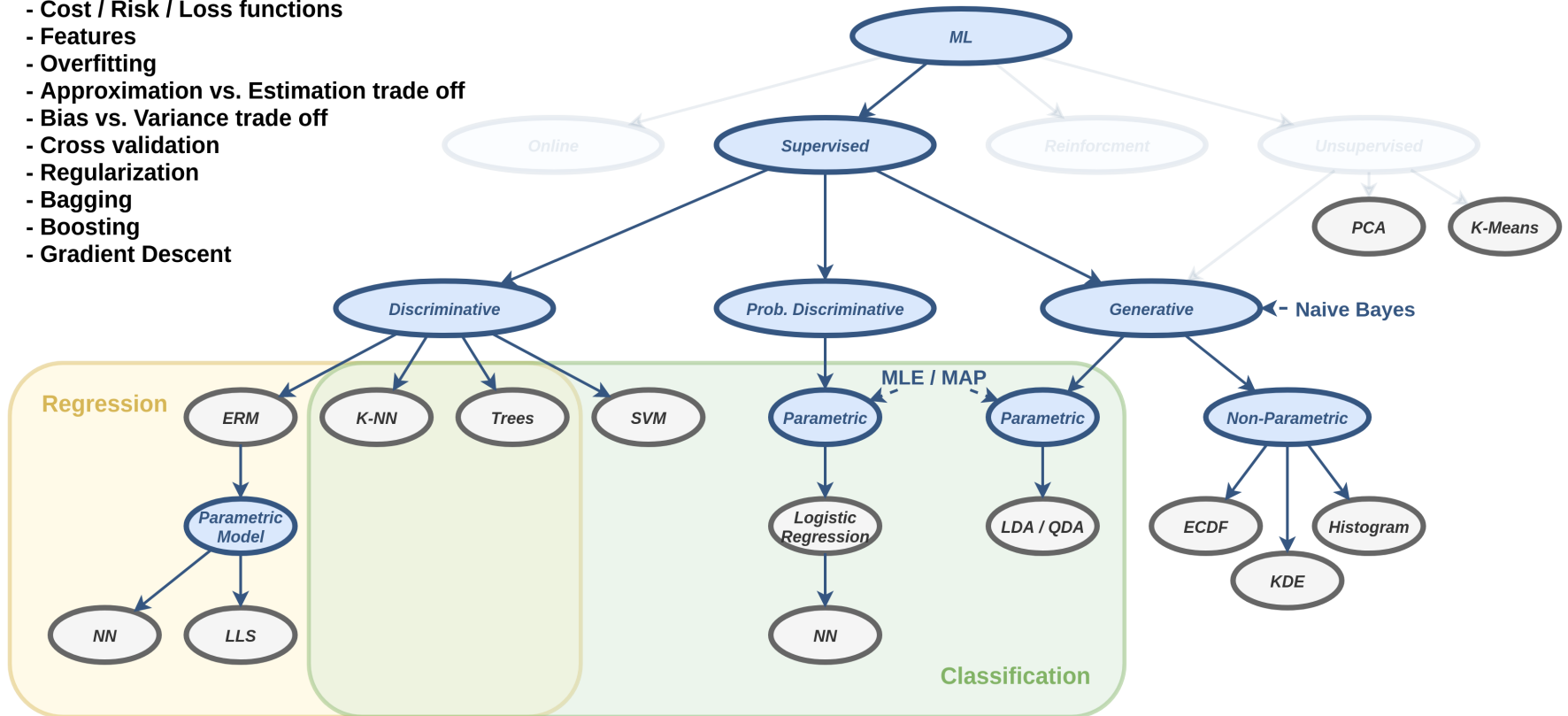
- בקורס נעסוק בעיקר בבעיות **Supervised learning** וניגע מעט ב**Unsupervised learning**.
- נכיר את עולם המושגים והשפה בה משתמשים בתחום.
- נלמד לבטא בעיות למידה באופן מתימטי.
- נכיר משפחות שונות של מודלים בהם ניתן השתמש.
- נלמד על שיטות לבחירת מודל "טוב" מתוך משפחת המודלים.
- נדון ביכולות והמגבלות של כל אחת מהשיטות והמודלים שנכיר.

- **13 הרצאות ותרגולים שבועיים.**
- **5 תרגלי בית יבש + רטוב. 15% מהציון.**
- **מבחן סופי 85% מהציון.**
- **בונוס: תרגילי הכנה של 0.2 נק' לכל תרגיל.**

Subjects Covered in this Course

General concepts:

- Cost / Risk / Loss functions
- Features
- Overfitting
- Approximation vs. Estimation trade off
- Bias vs. Variance trade off
- Cross validation
- Regularization
- Bagging
- Boosting
- Gradient Descent



בקורס נשתדל מאד להצמד נצמד לנוטציות המתמטיות
המופיעות בספר:

**Deep Learning (by I. Goodfellow, Y. Bengio & A.
Courville).**

את רשימת הנוטציות המלאה ניתן למצוא **קישור הבא**.

בשני התרגולים הראשונים יופיעו הנוטציות הקשורות
לאגברה לינארית והסתברות.

CSRankings: Computer Science Rankings

All Areas [\[off\]](#) [\[on\]](#)

AI [\[off\]](#) [\[on\]](#)

- ▶ Artificial intelligence
- ▶ Computer vision
- ▶ Machine learning & data mining
- ▶ Natural language processing
- ▶ The Web & information retrieval

#	Institution	Count	Faculty
1	▶ Carnegie Mellon University	152.8	55
2	▶ University of California - Berkeley	110.6	41
3	▶ Stanford University	94.2	40
4	▶ Massachusetts Institute of Technology	82.0	50
5	▶ Tsinghua University	69.4	51
6	▶ Cornell University	68.4	30
7	▶ Technion	66.9	27
8	▶ Univ. of Illinois at Urbana-Champaign	57.5	26
9	▶ University of California - Los Angeles	56.5	16
10	▶ University of California - San Diego	53.6	31

CSRanking (2010-2020)

Supervised learning

Supervised learning (למידה מונחית)

- בעיות supervised learning הן הבסיסיות ביותר בתחום והבנה טובה של בעיות אלו היא הבסיס להבנה של כל שאר הבעיות במערכות לומדות.
- בקורס זה אנו נעסוק בעיקר בבעיות מסוג זה.
- על מנת להבין מה הן בעיות supervised learning עלינו ראשית לחזור על הנושא של בעיות חיזוי.

- בבעיות חיזוי אנו מנסים לחזות את ערכו של משתנה אקראי לא ידוע, לרוב על סמך משתנים אקראיים ידועים.
- בעיות חיזוי הן **מאד** נפוצות ומופיעות במגוון רחב של תחומים בהנדסה ומדע.
- בהנדסת חשמל בעיות אלו מופיעות בתחומים כגון עיבוד אותות, תקשורת ספרתית ובקרה.
- בעיות חיזוי מלוות אותנו כמעט בכל פעולה יום יומית. לדוגמא האם לקחת מטריה כשיוצאים מהבית.
- ביום יום אנחנו לא מנסים לפתור בעיות אלה באופן מתמטי. אנו מחזיקים מודל של הקשרים הסטטיסטיים ומשתמשים בו בצורה איכותית.

הקשר ל supervised learning

- בבעיות חיזוי קלאסיות, אנו מניחים שהפילוג ידוע.

- ב supervised learning (ובמערכות לומדות) אנו מניחים כי הפילוג אינו ידוע.

- במקום הפילוג יש לנו מדגם.

- את החזאי נאלץ כעת לבנות על סמך המדגם (במקום על סמך הפילוג).

- **Labels** (תויות / תגיות): y - המשתנה האקראי שאותו אנו מנסים לחזות. (לרוב סקלר)
- **Observations \ measurements** (תצפיות או מדידות): x - הוקטור האקראי אשר שעלפיו נרצה לבצע את החיזוי. (לרוב וקטור)
- \hat{y} - תוצאת חיזוי.
- $\hat{y} = h(x)$ - פונקציית החיזוי.
- D אורך של הוקטור x

(המדגם) The dataset

המדגם יהיה מורכב מזוגות של x ו y אשר יוצרו מתוך N דגימות **בלתי תלויות**:

$$\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^N$$

N - מספר הדגימות שבמדגם.

- כל פונקציה אשר ממפה מ x ל y היא פונקציית חיזוי חוקית.
- היינו מעוניינים למצוא חזאי אשר לעולם לא טועה.
- מכיוון ש y משתנה אקראי לא נוכל לחזותו במדוייק.
- אנו צריכים להגדיר דרך להשוות בין הטעויות שאותם מבצעים החזאים שונים. (לדוגמא, הרבה טעויות קטנות או מעט גדולות)

Regression vs. Classification

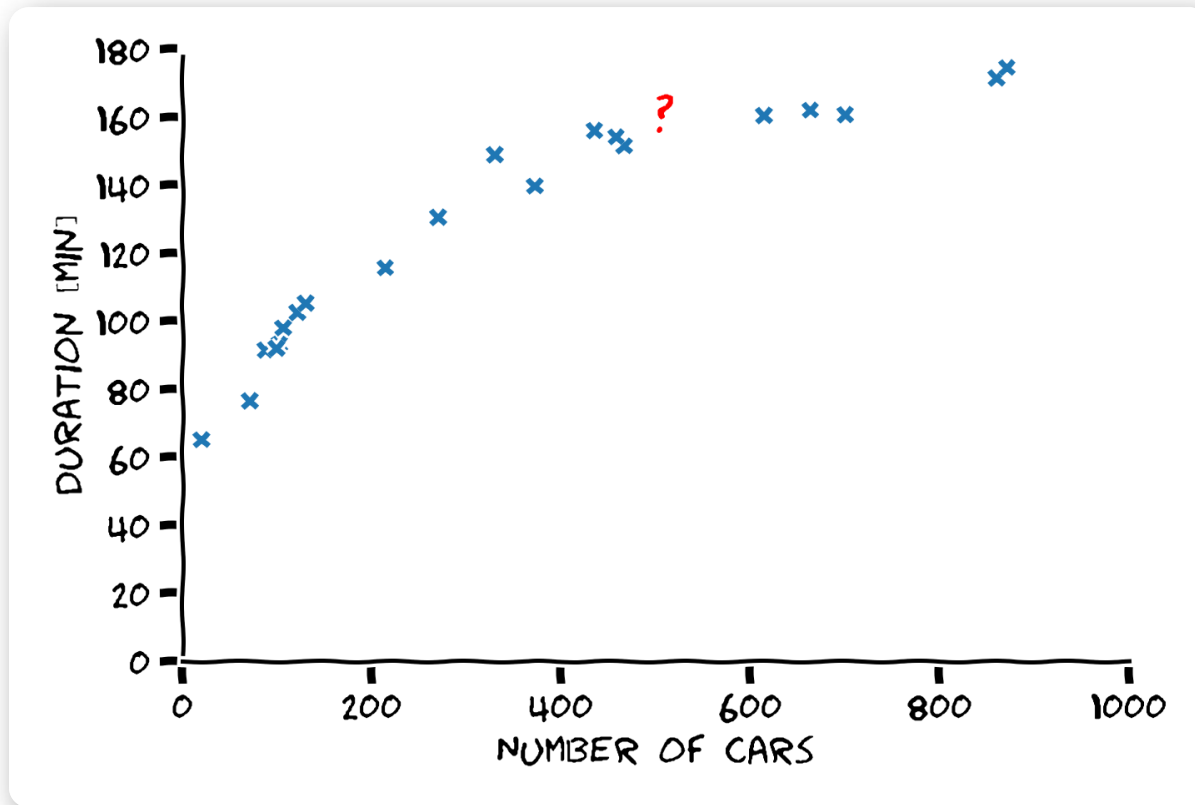
מוקבל לחלק את הבעיות ב supervised learning לשני תתי תחומים:

- בעיות regression (רגרסיה) - y רציף.

- בעיות classification (סיווג) - y בדיד עם סט ערכים סופי (לרוב קטן).

דוגמא לבעיית רגרסיה

הבעיה של חיזוי משך הנסיעה



ניסוח פורמלי - חיזוי משך הנסיעה

• x - מספר המכוניות על הכביש באותה נסיעה.

• y - משך הזמן שלקחה הנסיעה.

• נחפש פונקציית חיזוי $\hat{y} = h(x)$.

דוגמא לפונקציית חיזוי פרמטרית:

$$h_{\theta}(x; \theta_1, \theta_2) = \theta_1 + \theta_2 x$$

אוסף כל הפונקציות הלינאריות (ליתר דיוק אפיניות).

בהמשך נעסוק בשאלה של כיצד לבחור את הערכים של θ_1 ו- θ_2 .

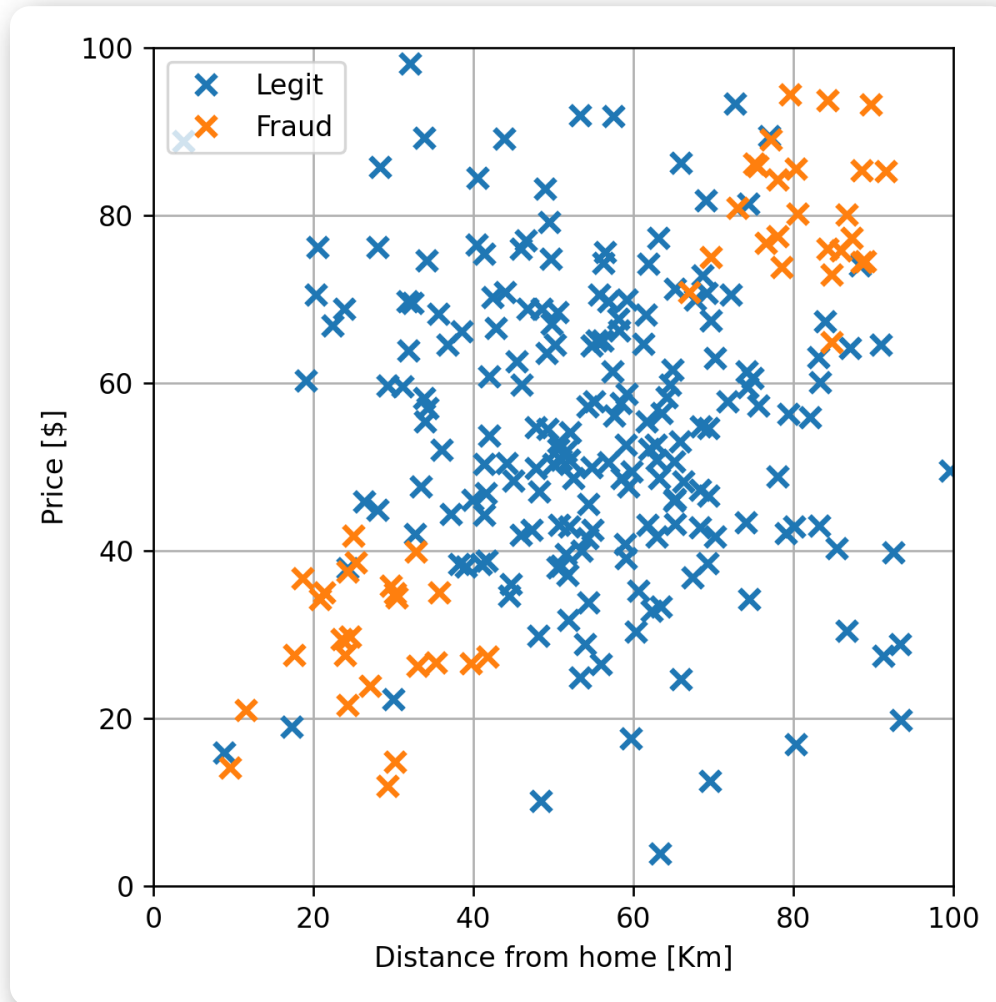
בעיית סיווג - זיהוי הונאות בכרטיסי אשראי

נסיון לסווג עסקאות כחשודות להונאה על סמך פרטי העסקה:

- הסכום.
- מרחק העיסקה (נניח המיקום של החנות) מהעיסקה האחרונה.
- מרחק העיסקה מהכתובת של הלקוח.
- השעה ביום.
- אופי המוצרים שהחנות מוכרת (מכולת, מוצרי חשמל, ביגוד, רכב, נדל"ן, וכו')
- ניתן לעשות זאת בעזרת supervised learning על ידי שימוש במדגם דוגמאות מהעבר.

בעיית סיווג לדוגמא - המדגם

בהרצאות נשתמש בבעיה זו כדוגמא ונתייחס למדגם הבא:



בעיית סיווג לדוגמא - החזאי

נרצה למצוא חזאי אשר יחזה עיסקאות חשודות על פי מרחק ומחיר. לדוגמא:

