

Core concepts in Data Analytics and ML

סוגי למידת מכונה

Supervised Learning
Unsupervised Learning
Reinforcement Learning

רגרסיה לינארית

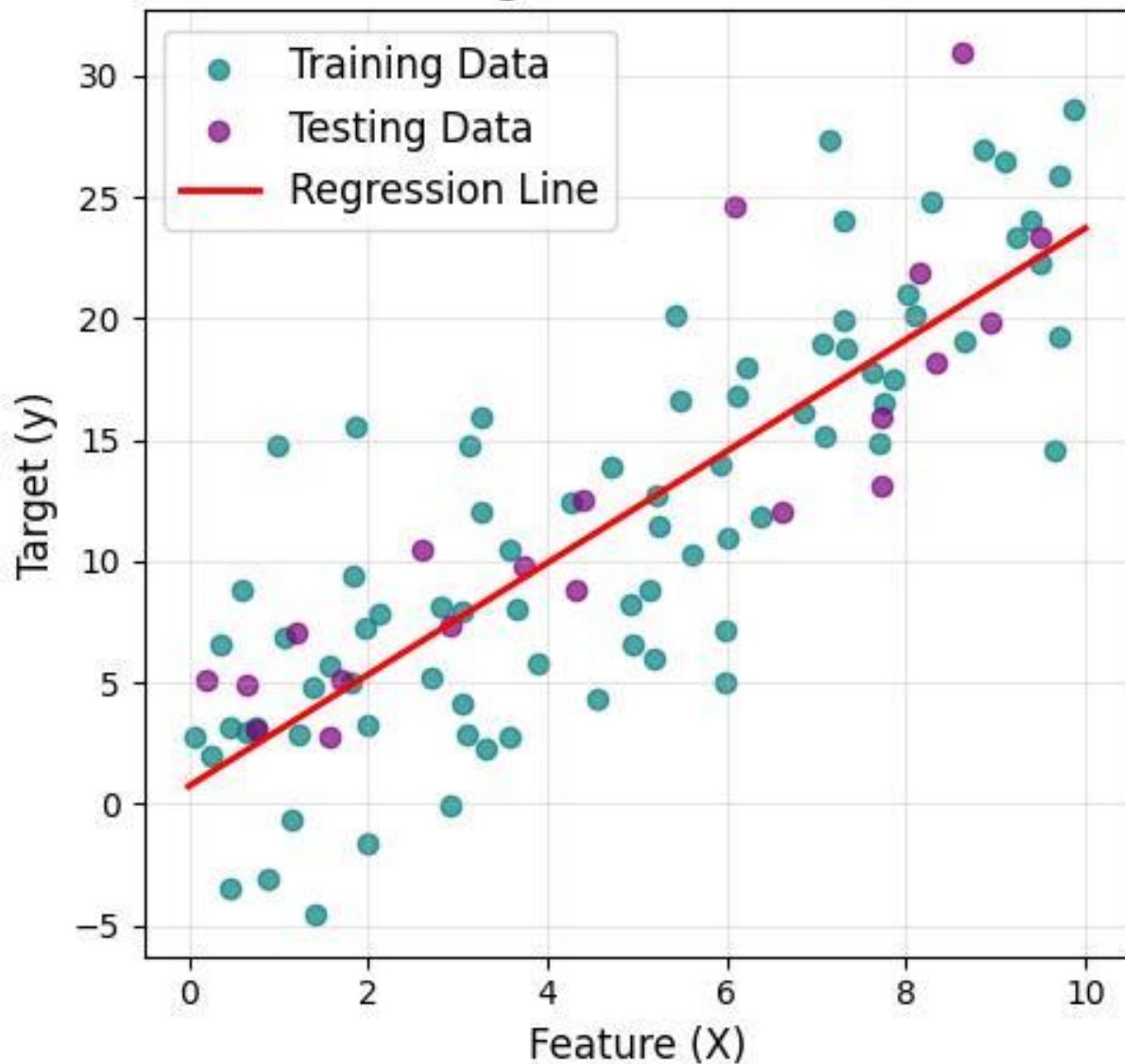
Features
Target/Label
prediction
Weight
bias





למידת מכונה היא תת-תחום של בינה
מלאכותית שבו מחשבים לומדים לזהות
דפוסים מנתונים, במקום להיות
מתוכנתים באופן מפורש.
מקום לתכנת את הכללים, אנו מלמדים
את המחשב לזהות אותם לבד.

Linear Regression Visualization



למידה מפוקחת: (Supervised Learning)

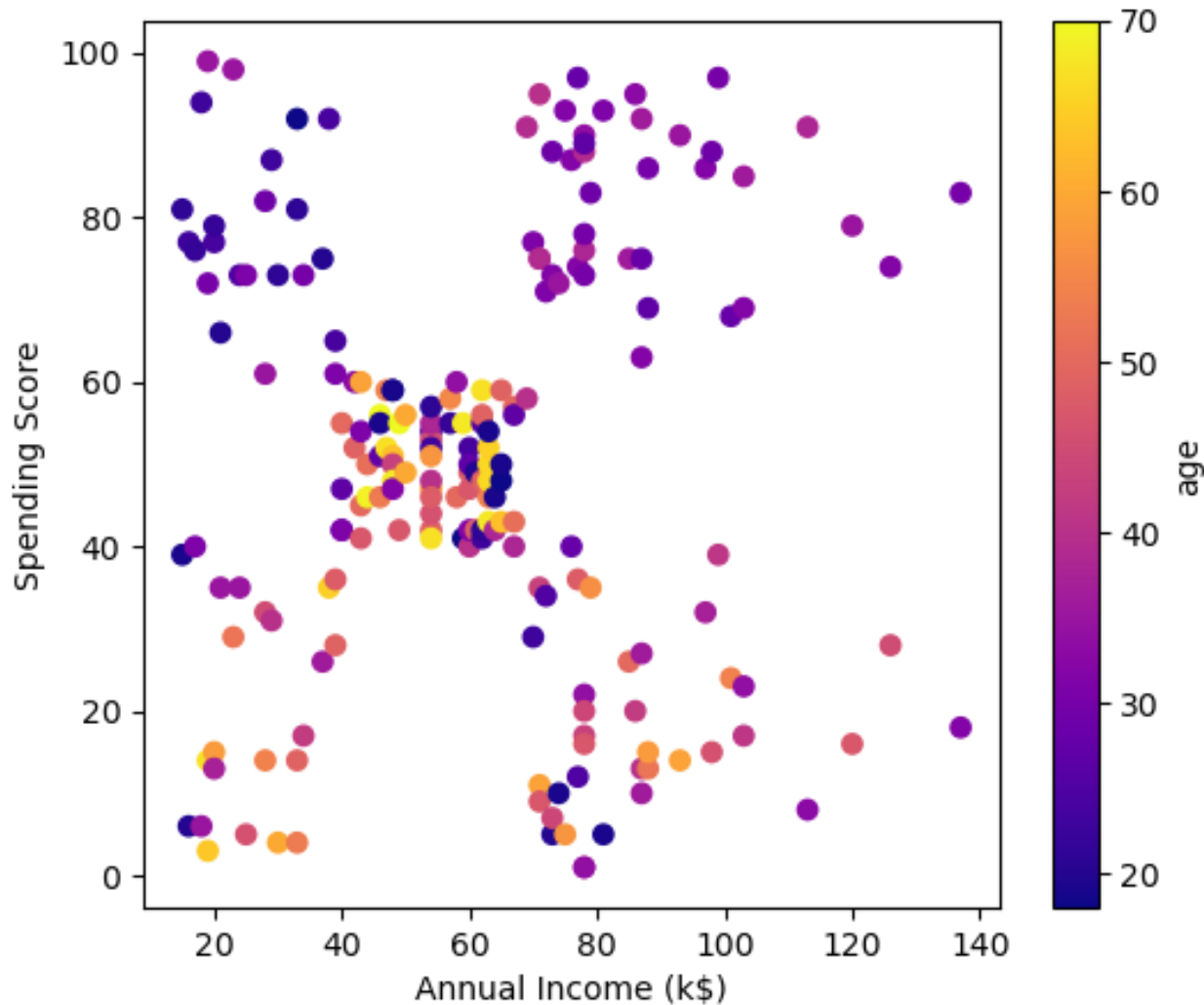
למידה על נתונים עם תגיות.

המחשב לומד על סמך נתונים "מתויגים" –
כלומר, יש תשובה נכונה לכל דוגמה באימון.

דוגמאות נפוצות:

חיזוי Regression – ניבוי ערכים כמותיים
סיווג – classification – חלוקה לקבוצות.

<https://teachablemachine.withgoogle.com/train>



• למידה לא-מפוקחת (Unsupervised Learning):

זיהוי דפוסים ללא תגיות.

• **Clustering:** סיווג לקבוצות לפי דמיון.

דוגמה: קיבוץ לקוחות לפי התנהגות רכישה.

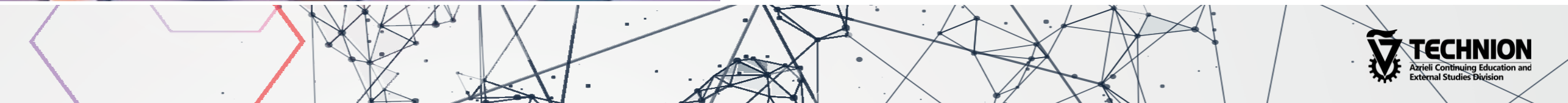
• **Association:** זיהוי קשרים בין פריטים.

• דוגמה: זיהוי שאנשים שקונים חיתולים נוטים

לקנות גם בירה.



למידת חיזוק (Reinforcement Learning)
למידה באמצעות ניסוי וטעייה עם תגמולים ועונשים.
דוגמה: alphazero



משימה	חיזוי regression	סיווג classification
זיהוי דואר זבל		זבל / לא זבל
מה מחיר דירה ע"פ שטח	ניבוי מחיר	
הגדרת מצב רפואי	רמת סיכון באחוזים	חלوه / בריא

Machine Learning

Supervised Learning

Unsupervised Learning

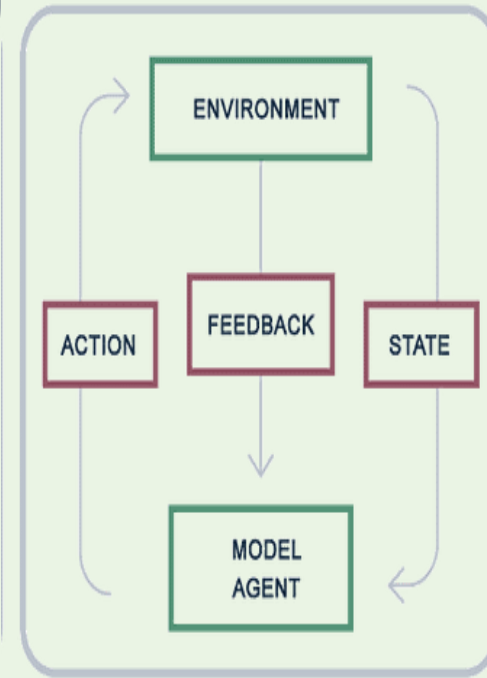
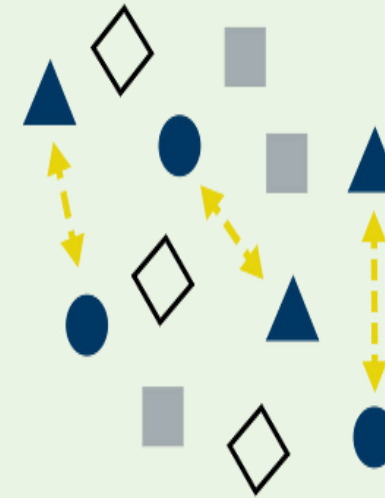
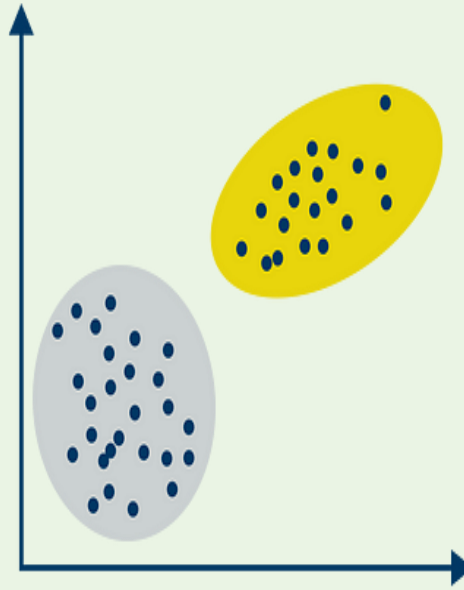
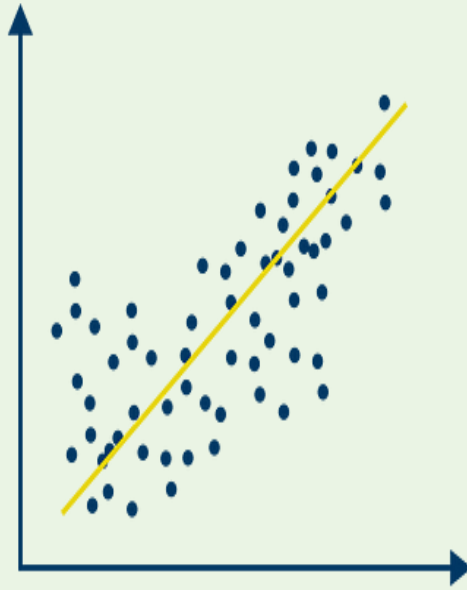
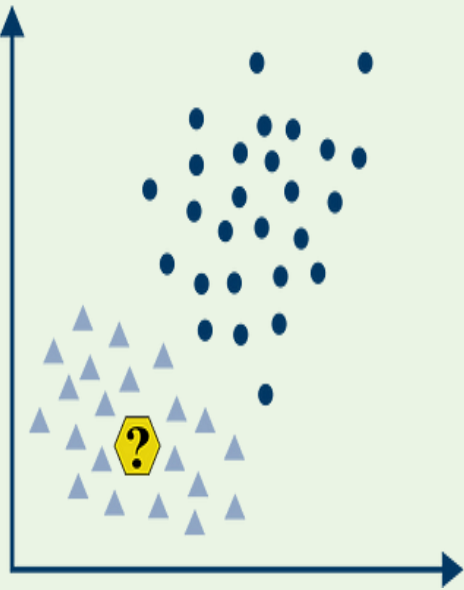
Reinforcement Learning

Classification

Regression

Clustering

Association

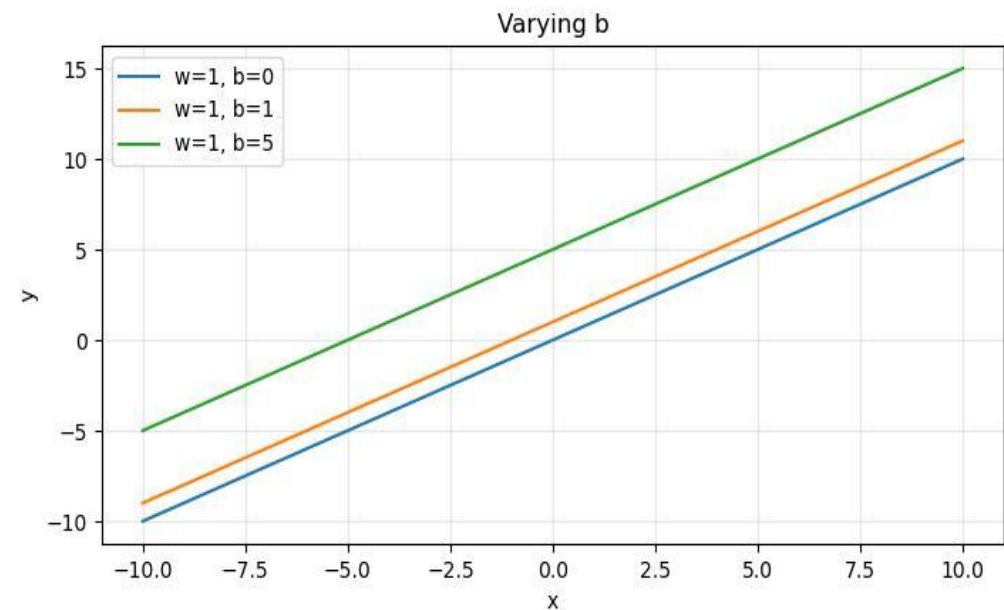
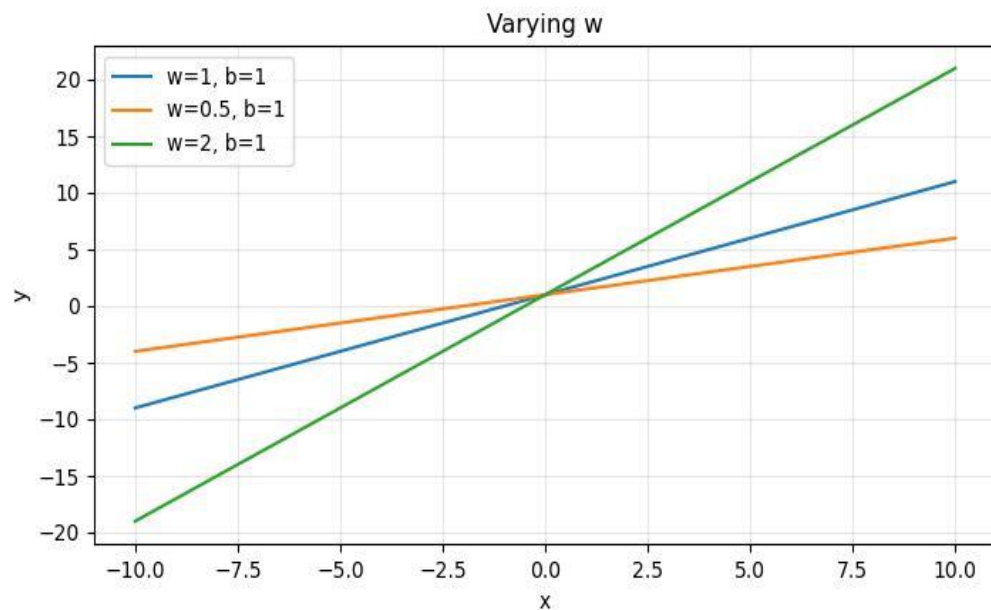




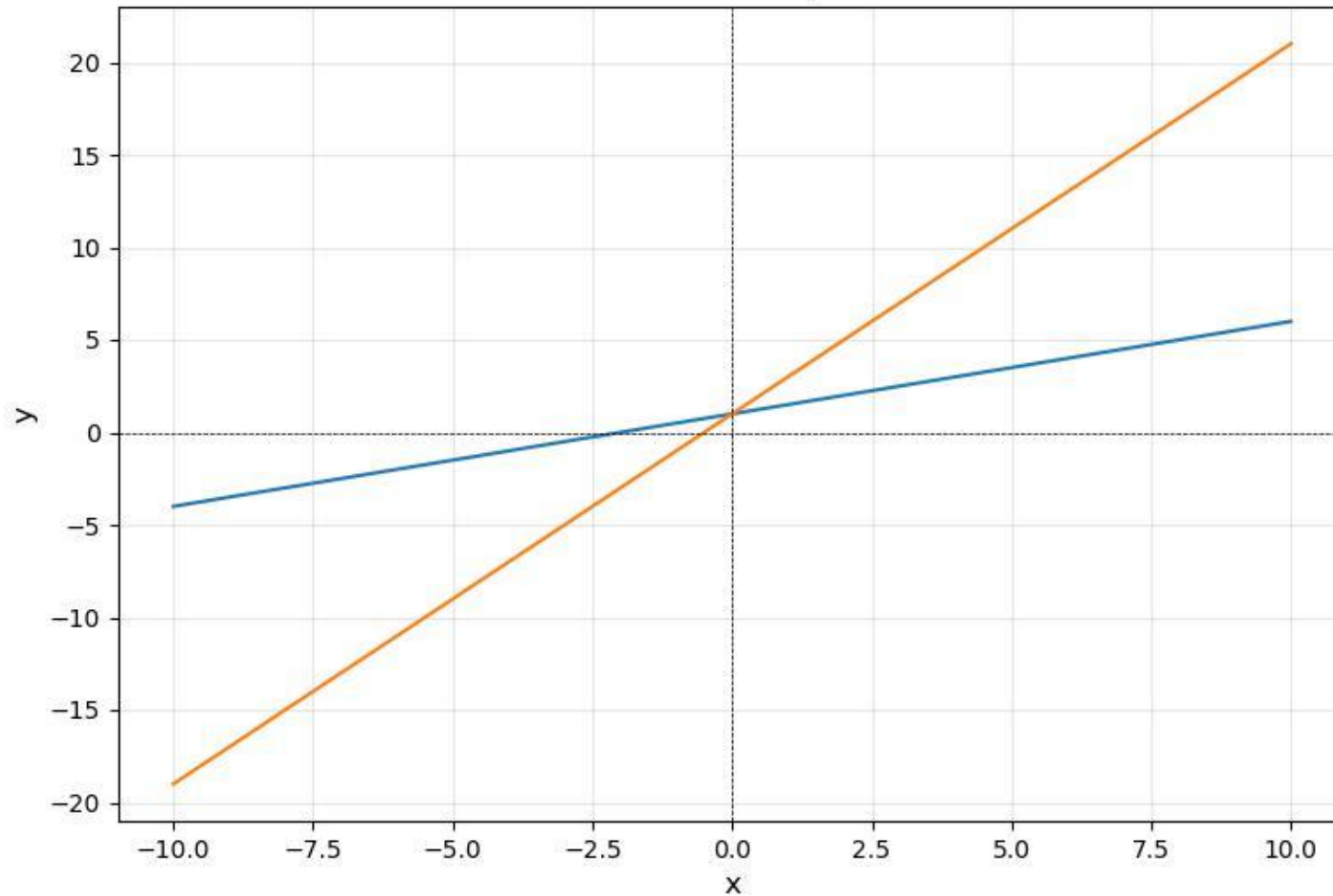
Linear regression

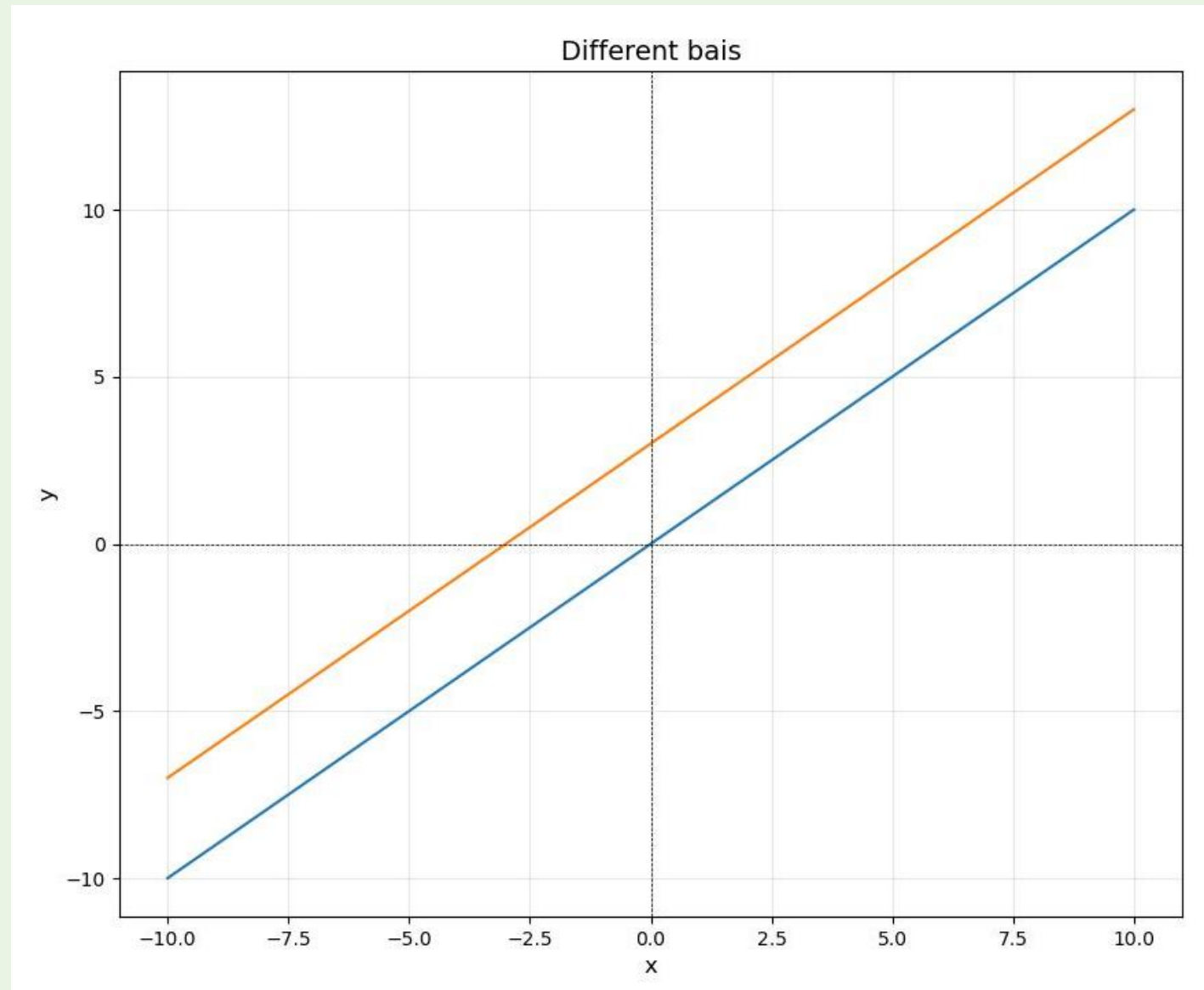
רגרסיה לינארית היא מודל שמנסה למצוא קו ישר שמייצג את הקשר בין משתנה תלוי למשתנה עצמאי (בלתי תלוי).
המשוואה:

$$y = wx + b$$



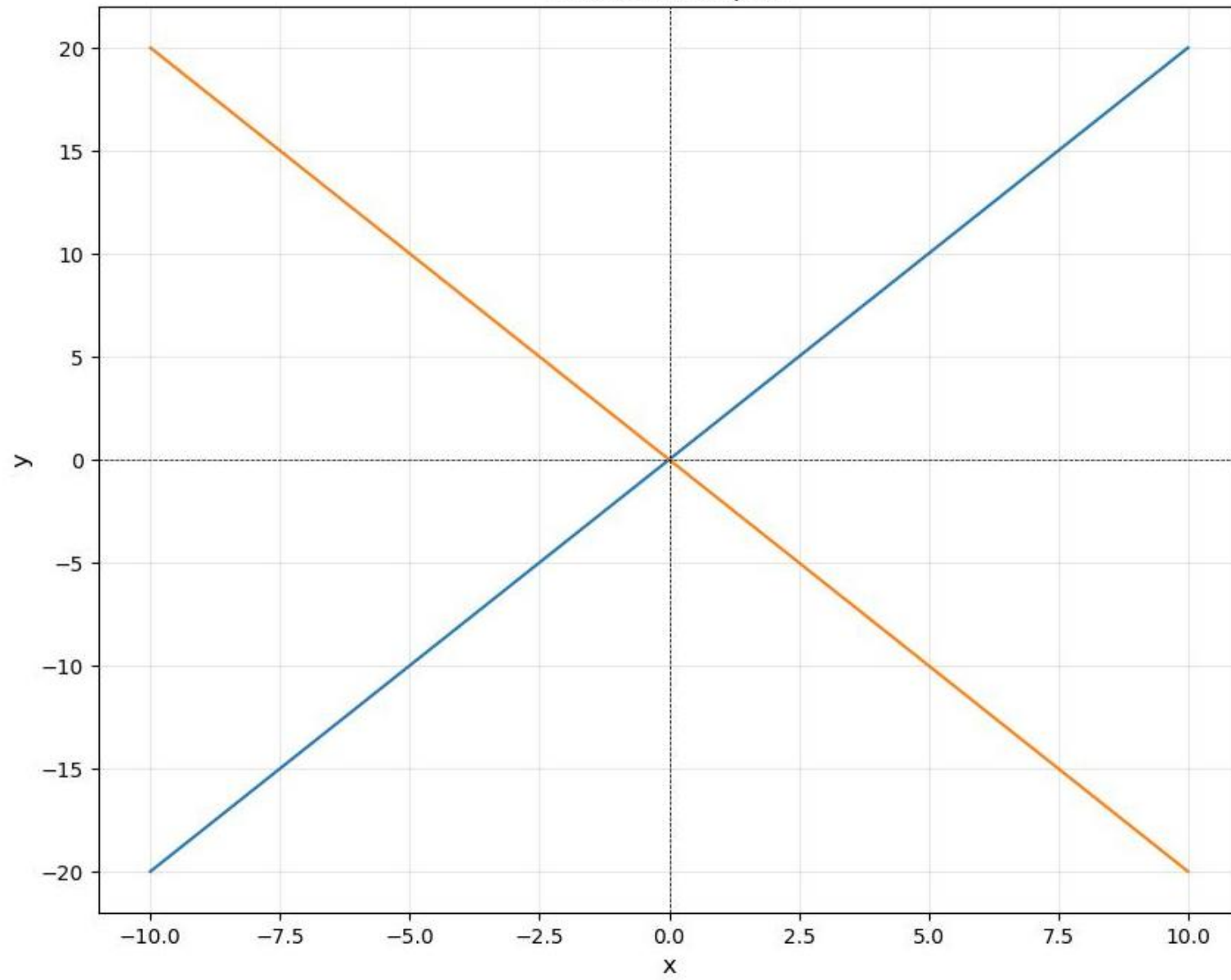
Different Slopes



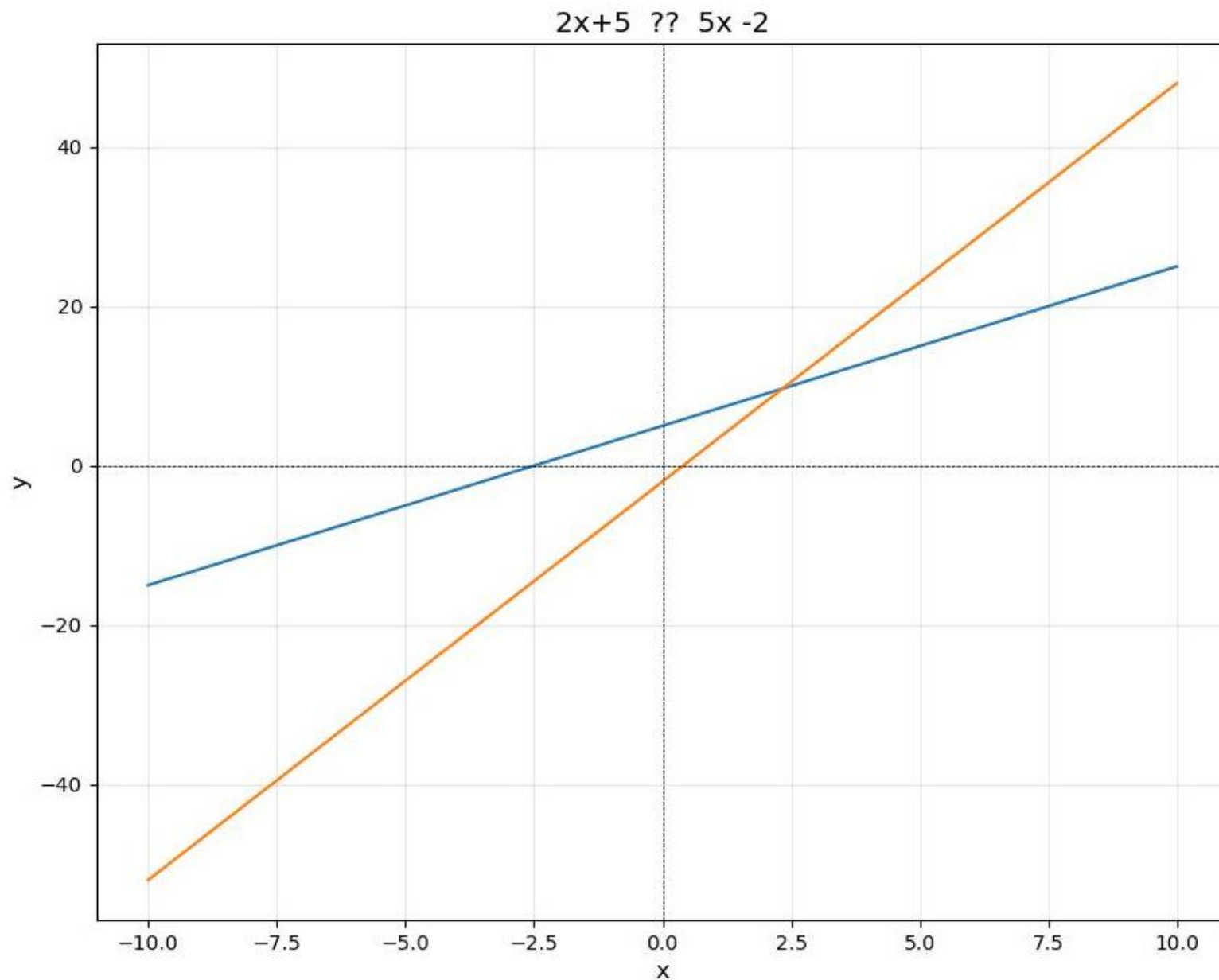


Linear regression

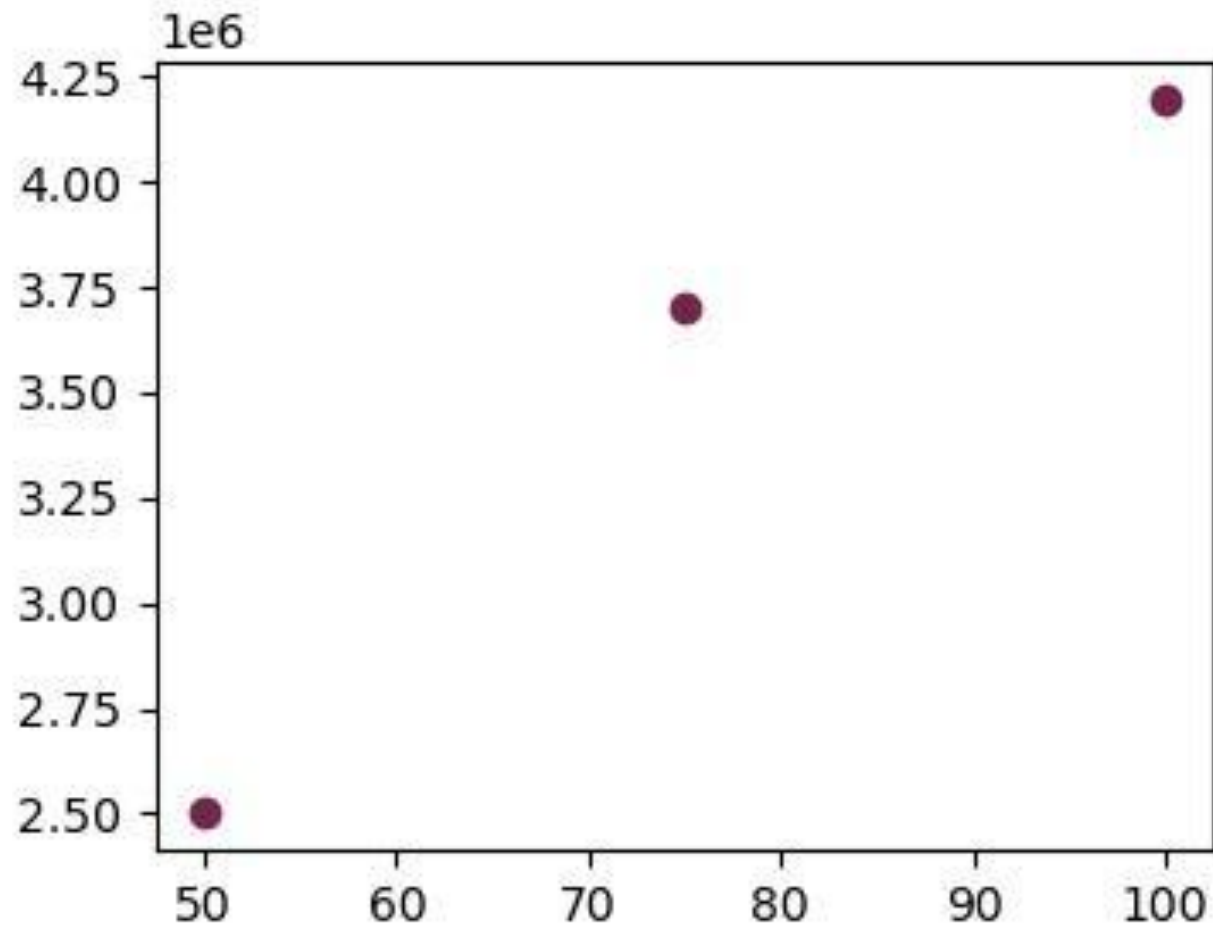
Different Slopes



Linear regression



Linear regression



מחיר מוגזם	גודל דירה
2500000	50
3700000	75
4200000	100

$$y = wx + b$$

X	Features
y	Target/Label
\hat{y}	prediction
$x^{(i)}, y^{(i)}$	The i^{th} training sample
m	Number of training samples
w	Weight/ confidence
b	bias

OLS – Ordinary Least Squares

שיטה סטטיסטית שמטרתה למצוא את הקו הישר (או המשטח) שמתאים לנתונים בצורה הטובה ביותר.

OLS עושה זאת על ידי מזעור סכום ריבועי השגיאות (Sum of Squared Errors, SSE)

$$\min_w \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

OLS מחפש את הערכים של w שממזערים את ריבועי ההבדלים במטרה למצוא את המקדמים הטובים ביותר למודל (למשל, בשיטת הרגרסיה הליניארית).

המודל המתקבל הוא המודל שבו MSE (ערך השגיאה) יהיה הקטן ביותר