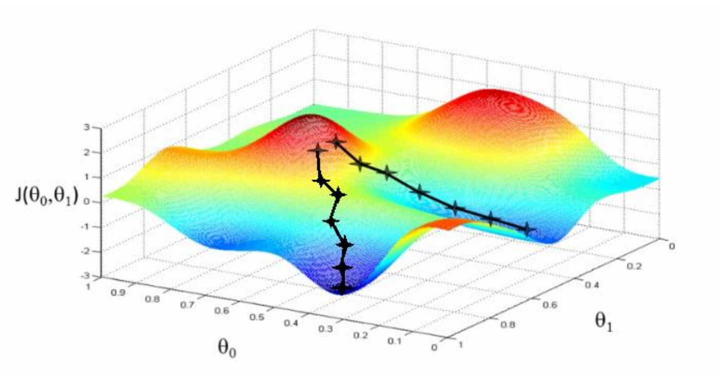
לא לשכוח להפעיל הקלטה!

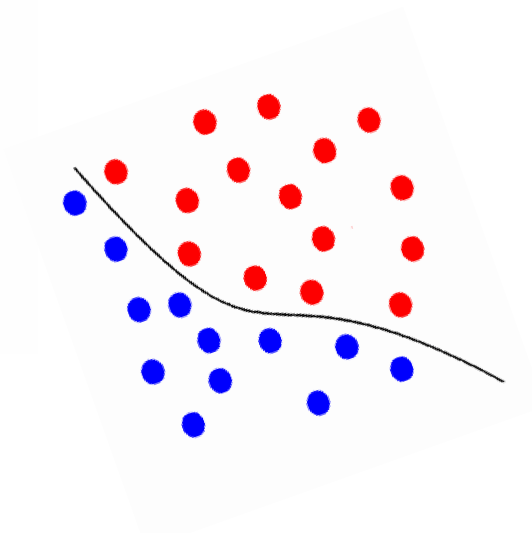
מה נלמד בשיעור זה?

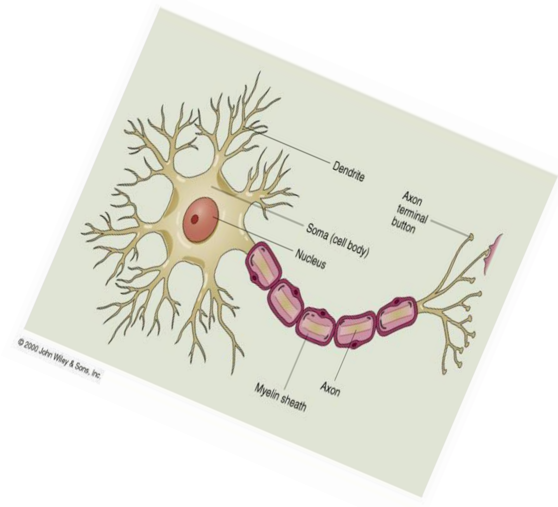
- ✓ רגרסיה לינארית עם משתנה יחיד
- ✓ gradient descent עבור רגרסיה ליניארית למשתנה יחיד
- ✓ רגרסיה לינארית עם ספריית scikit-learn
- ✓ הערכת ביצועים r^2

מכון דוידסון לחינוך מדעי
DAVIDSON INSTITUTE OF SCIENCE EDUCATION
מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

רגרסיה ליניארית



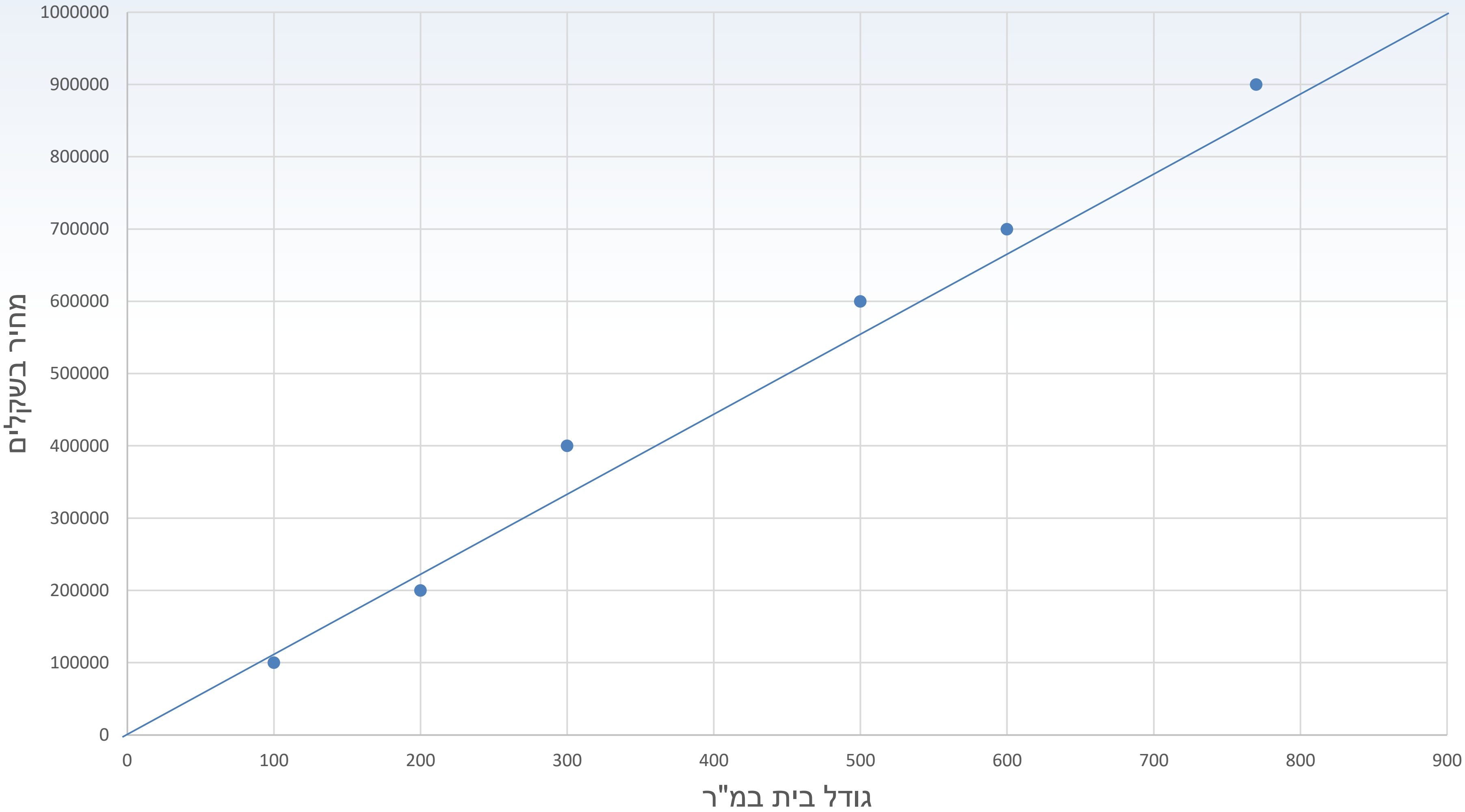




שיעור סינכרוני

רגרסיה לינארית עם משתנה יחיד

רגרסיה לינארית



מטרות מציאת מודל?

מדוע העברתם את הקו? למה לא להסתפק רק בהצגת הנתונים?

- ✓ דרך להעריך מתמטית מה עומד מאחורי הנתונים (למשל, אם אלו נתוני מתח-זרם, נוכל להסיק משיפוע הקו את ההתנגדות).
- ✓ data compression - דרך להציג את הנתונים בצורה הפשוטה שלהם. במקום להציג את כולם – להציג רק נוסחא.
- ✓ דרך לאפשר יכולת של חיזוי על נקודות שאינן חלק מהנתונים

מהי רגרסיה?

- מציאת קו המתאר את הנתונים
- בבעיית רגרסיה, בניגוד לבעיית קלסיפיקציה, אנחנו מנסים למצוא פונקציה מסוימת המקשרת את סט הנתונים לפלט מסוים.
- מבחינת ערכי הפלט (y) – אנו מנסים לחזות את **הערך המדויק של y**, לא רק לחלק את הנתונים למחלקות (הפלט הוא לא רק 0 או 1).

חסרונות ויתרונות של רגרסיה ליניארית

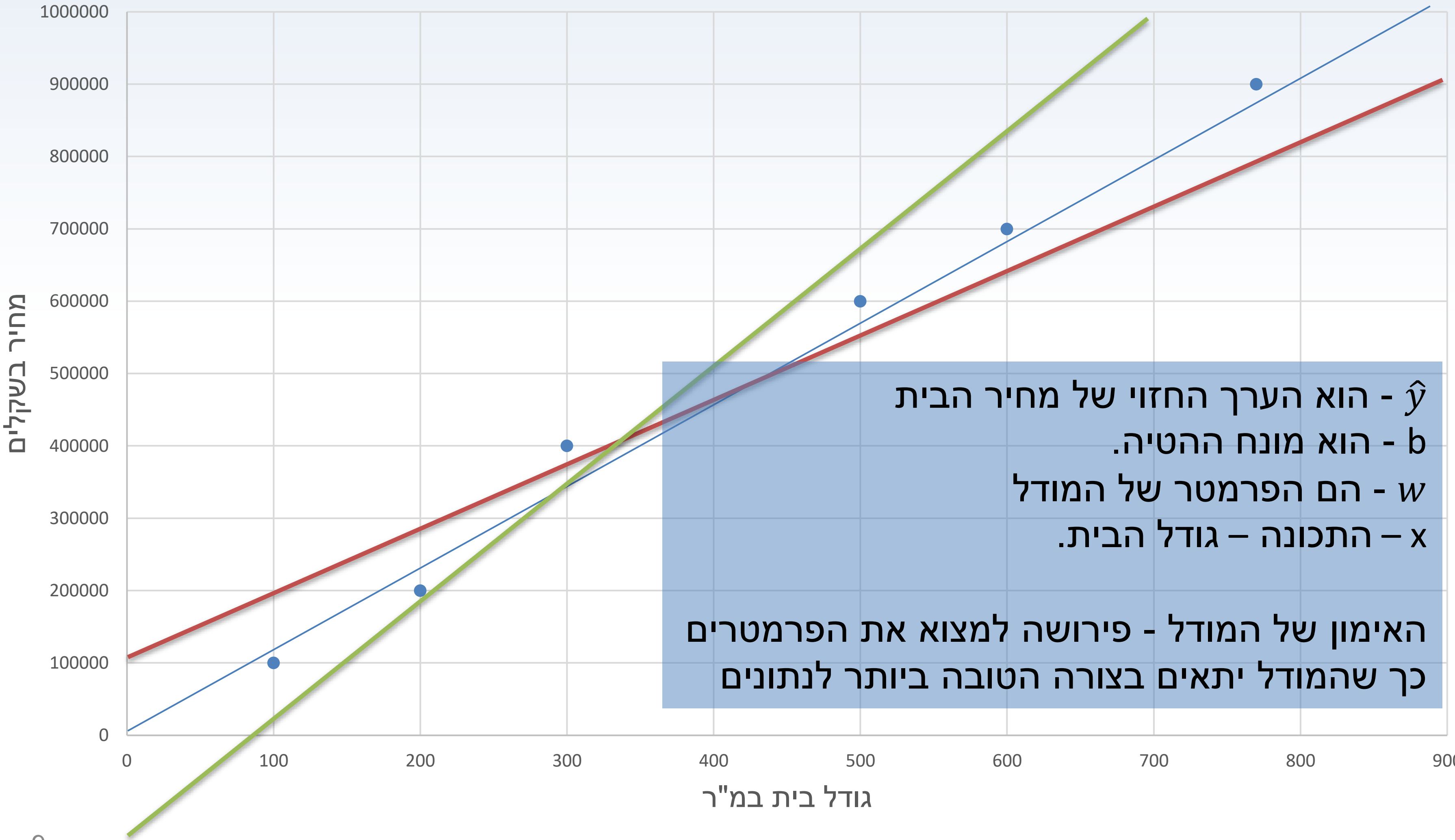
מודל המניח יחסים ליניארים בין המשתנים: $\hat{y} = w \cdot x + b$

מודל ליניארי ורגרסיה ליניארית יהיו טובים במקרה שיש קשר ליניארי בין המשתנים.
אלו מקרים כאלו אתם מכירים?
זרם לעומת מתח,
גובה הורים לעומת גובה ילדים,
הכנסה לעומת השכלה וכו'

ישנם מקרים שבהם יש קשר בין המשתנים, אבל הקשר לא לינארי:
במקרים האלה נבצע רגרסיה לא ליניארית.

כיצד בוחרים מודל?

$$\hat{y} = w \cdot x + b$$



$\hat{y}$ - הוא הערך החזוי של מחיר הבית
b - הוא מונח ההטיה.
w - הם הפרמטר של המודל
x – התכונה – גודל הבית.

האימון של המודל - פירושה למצוא את הפרמטרים
כך שהמודל יתאים בצורה הטובה ביותר לנתונים

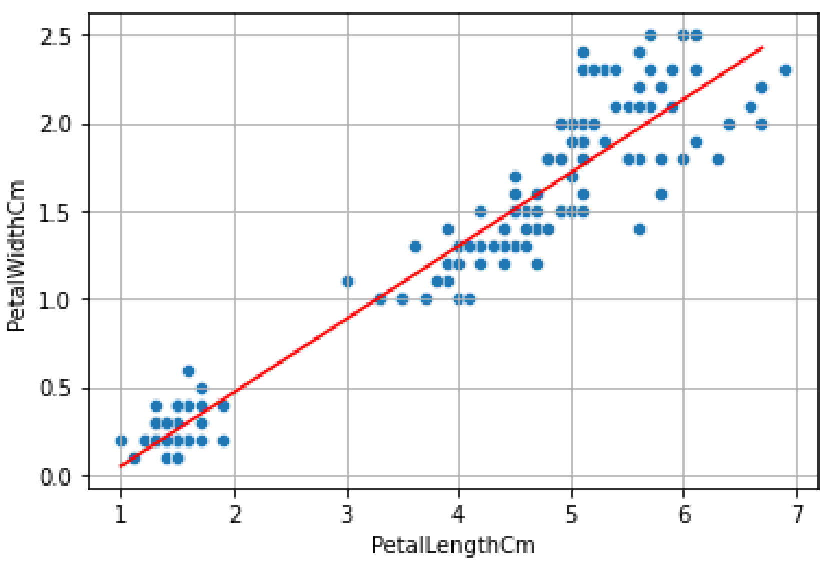
תרגיל 1 – חיזוי על ידי רגרסיה לינארית

ברגרסיה לינארית לחיזוי רוחב עלה של אירוס על בסיס אורכו נמצאו הפרמטרים הבאים:

- w=0.416
- b=-0.366

מה יהיה חיזוי רוחב העלה עבור הפרחים הבאים:

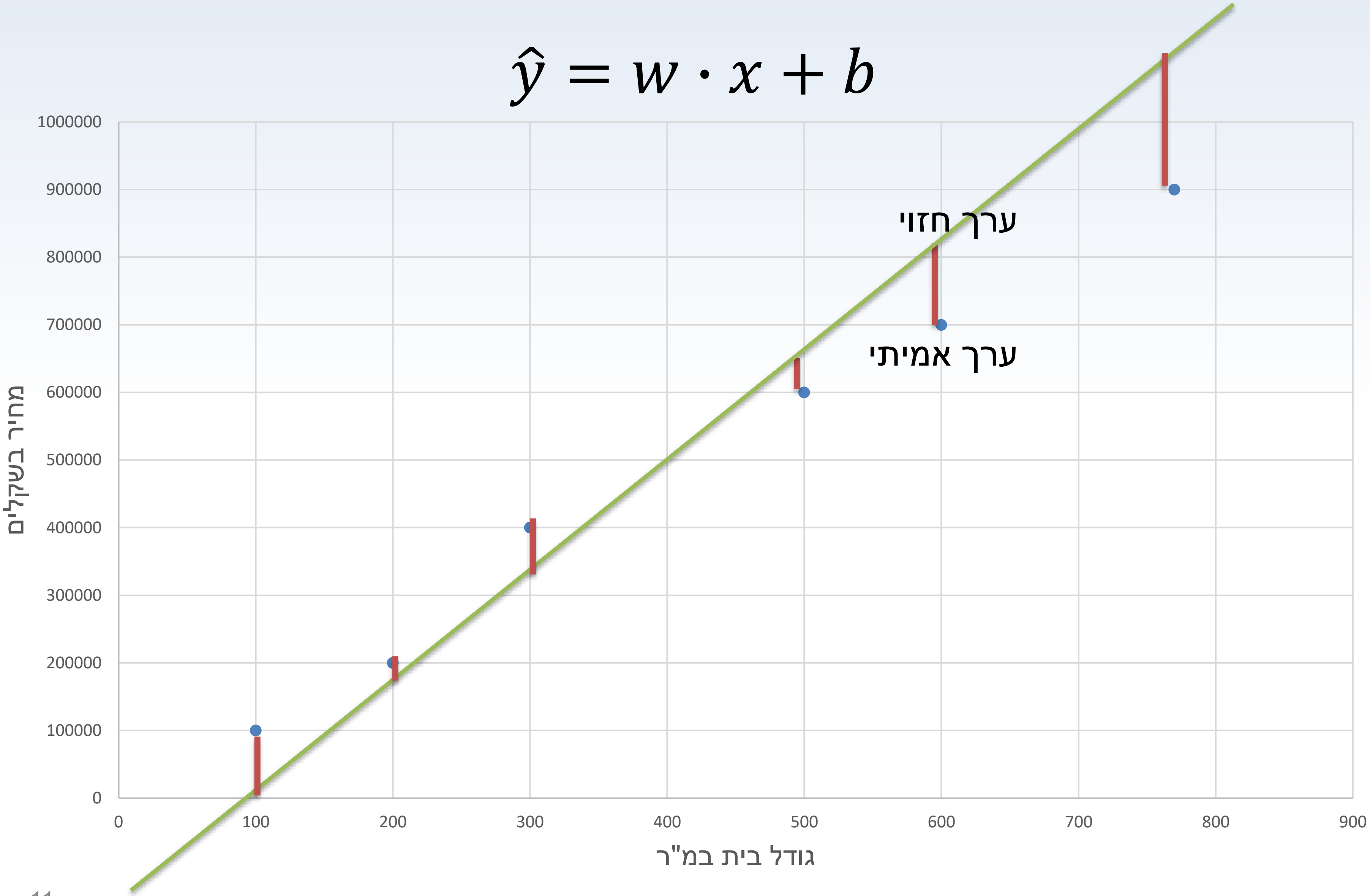
$$\hat{y} = w \cdot x + b$$



Petal length	Petal width
1.5	
2.5	
6.5	
0.5	
0	

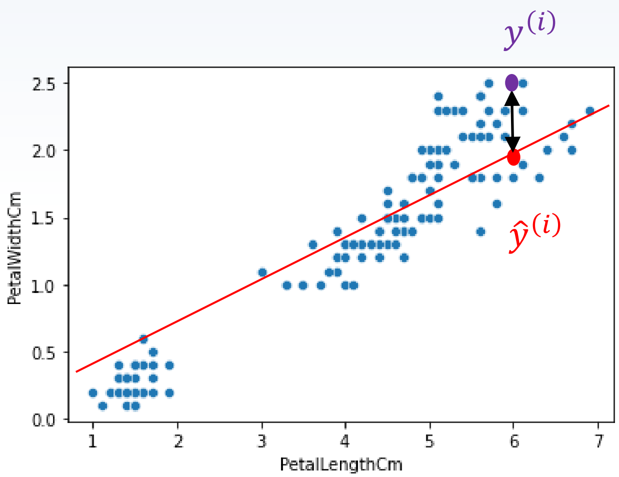
איך נחליט האם המודל טוב?

$$\hat{y} = w \cdot x + b$$



שגיאת ריבועית על הדוגמא ה-(i)

- בהינתן סט של m דוגמאות $(x^{(i)},y^{(i)})$
  - $x^{(i)}$ - ערכי דגימות של המשתנה הבלתי תלוי
  - $y^{(i)}$ – ערכי דגימות המשתנה התלוי בהתאמה



- ובהינתן מודל רגרסיה $\hat{y} = w \cdot x + b$

- נגדיר את שגיאת החיזוי על הדוגמא ה-(i):

$$l^{(i)} = (\hat{y}^{(i)} - y^{(i)})^2 = (w \cdot x^{(i)} + b - y^{(i)})^2$$

פונקציית המחיר J

- בהינתן סט של m דוגמאות $(x^{(i)},y^{(i)})$
  - x_i - ערכי דגימות של המשתנה הבלתי תלוי
  - y_i – ערכי דגימות המשתנה התלוי בהתאמה

- ובהינתן מודל רגרסיה $\hat{y} = w \cdot x + b$
- נגדיר את שגיאת החיזוי על כל סט נתוני האימון:

$$J = \frac{1}{m}\sum_{i=1}^{m} l^{(i)} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{m}\sum_{i=1}^{m}(w \cdot x^{(i)} + b - y^{(i)})^2$$

פונקציית המחיר J היא פונקציה של W,B

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}(\underbrace{w \cdot x^{(i)} + b}_{\text{הערך החזוי}} - \underbrace{y^{(i)}}_{\text{הערך האמיתי}})^2$$

<u>המטרה</u> – למזער את הטעות עבור ערכי w,b שונים

- w*x+b תשובות המודל
- y הנתונים האמיתיים
- m – מספר הדגימות
- i – אינדקס שרץ על כל הדגימות

Cost Function

$$J(w,b)=\frac{1}{m}\sum_{i=1}^{m}(\underbrace{w\cdot x^{(i)}+b}_{\text{הערך החזוי } \hat{y}}-\underbrace{y^{(i)}}_{\text{הערך האמיתי}})^2$$

MSE (**M**EAN **S**QUARED **E**RROR)

רגרסיה לינארית עם gradient descent

אלגוריתם GRADIENT DESCENT

- על מנת למצוא את נקודת המינימום של פונקציה J(w):
  - נגריל נקודת התחלה w_0
  - נקדם את ערך w לפי הנוסחה:

$$w \leftarrow w - \alpha \frac{\partial J(w)}{\partial w}$$

- α – קצב הלמידה
- $\frac{\partial J(w)}{\partial w}$ - נגזרת הפונקציה

אלגוריתם GRADIENT DESCENT עבור פונקציה רב ממדית

- על מנת למצוא את נקודת המינימום של פונקציה J(w,b):
  - נגריל נקודת התחלה w_0,b_0
  - נקדם את ערכי w,b לפי הנוסחאות:

$$w \leftarrow w - \alpha \frac{\partial J(w,b)}{\partial w}$$

$$b \leftarrow b - \alpha \frac{\partial J(w,b)}{\partial b}$$

שלבים למציאת המשקולות באמצעות gradient descent

- ✓ 1. איתחול W, b - ראשית אנו מאתחלים את פרמטרי המודל בערכים רנדומלים
- ✓ 2. חישוב הנגזרת החלקית ל - W, b
- ✓ 3. עידכון b – הערך החדש של b שווה לערך הקודם של b פחות הנגזרת של b כפול גודל הצעד
- ✓ 4. עידכון w – הערך החדש של w שווה לערך הקודם של w פחות הנגזרת של w כפול גודל הצעד
- ✓ 5. בדיקת ביצועים – חישוב r^2

אנו חוזרים על שלבים 2-4 עד שפונקציית העלות תתכנס לערך המינימלי

פונקציית המחיר ונגזרותיה

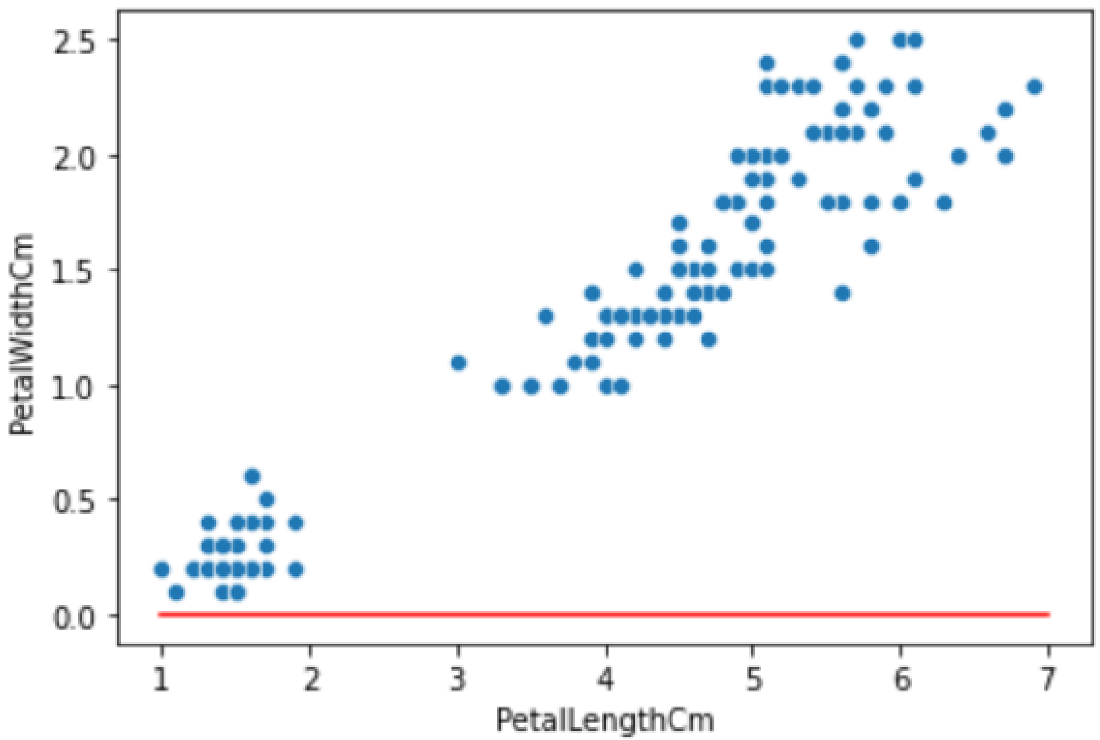
$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}(w \cdot x^{(i)} + b - y^{(i)})^2$$

$$\frac{\partial}{\partial w}J(w,b) = \frac{1}{m}\sum_{i=1}^{m}2\left(w \cdot x^{(i)} + b - y^{(i)}\right)x^{(i)}$$

$$\frac{\partial}{\partial b}J(w,b) = \frac{1}{m}\sum_{i=1}^{m}2\left(w \cdot x^{(i)} + b - y^{(i)}\right)$$

אתחול

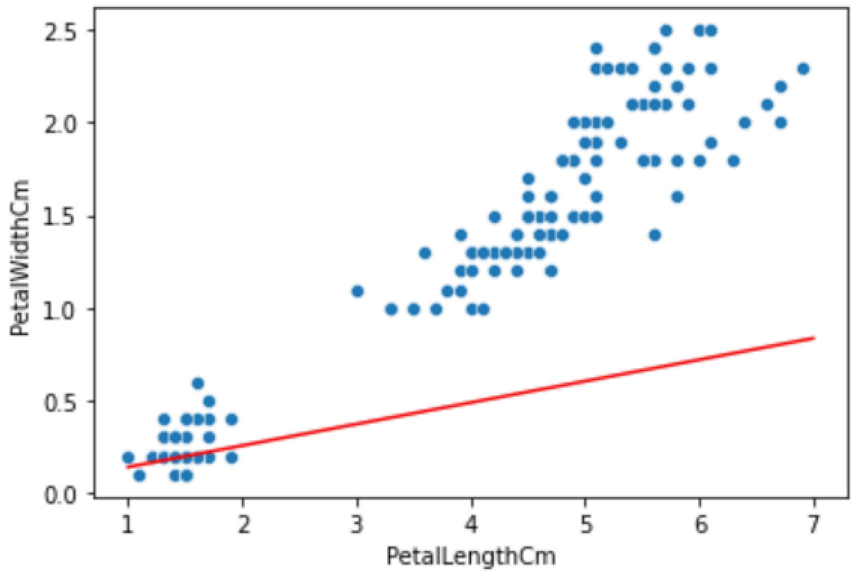
```
w = 0
b = 0
alpha = 0.01

plot_reg_line(w,b)
```



צעד למידה

```
dw_ = dw(w,b)
db_ = db(w,b)

w = w - alpha * dw_
b = b - alpha * db_

plot_reg_line(w,b)
print (w,b,J(w,b))
```

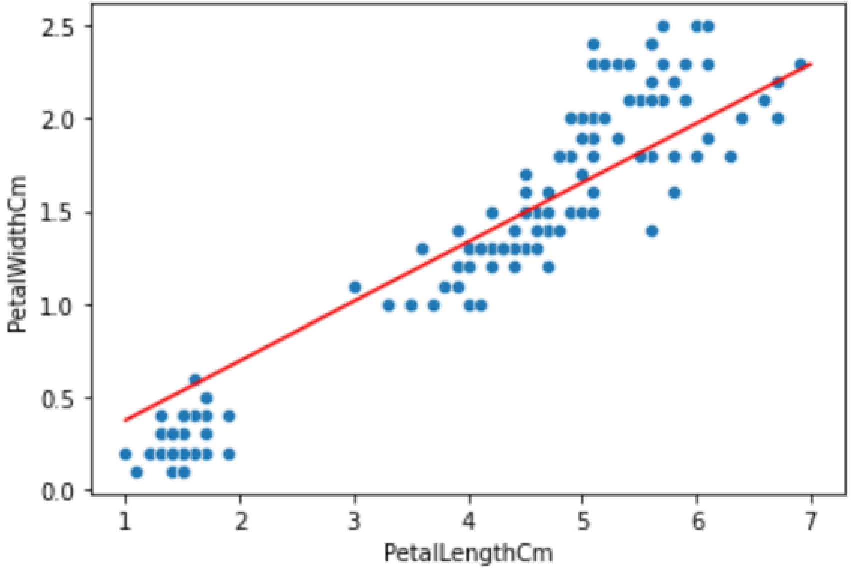
```
0.11586266666666667 0.023973333333333336 0.8680644756207763
```



אחרי 10 צעדים

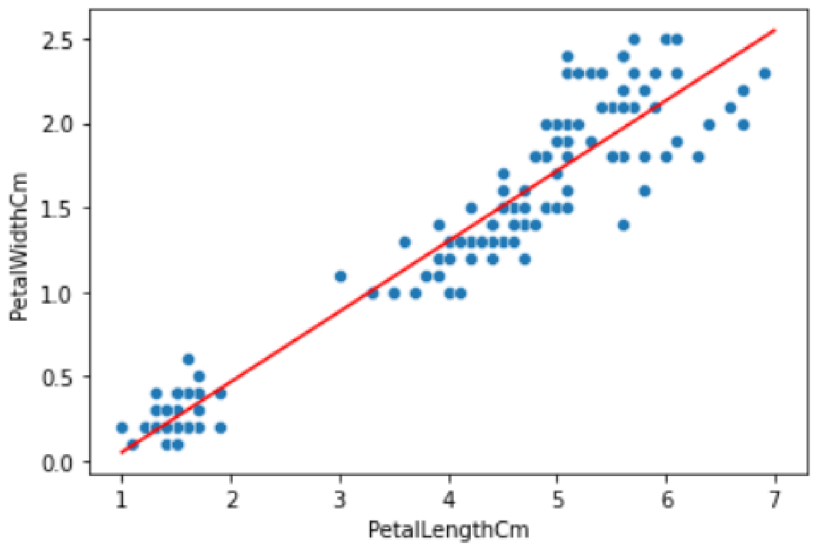
```
dw_ = dw(w,b)
db_ = db(w,b)

w = w - alpha * dw_
b = b - alpha * db_

plot_reg_line(w,b)
print (w,b,J(w,b))
```

```
0.3196755592709454 0.05501231875098751 0.07458529160580048
```



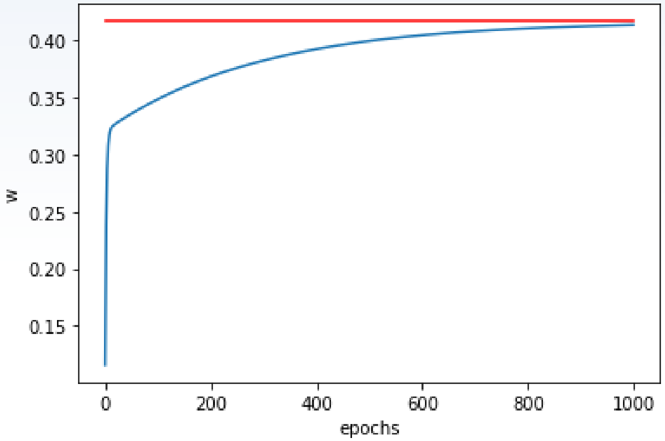
אחרי 1000 צעדים

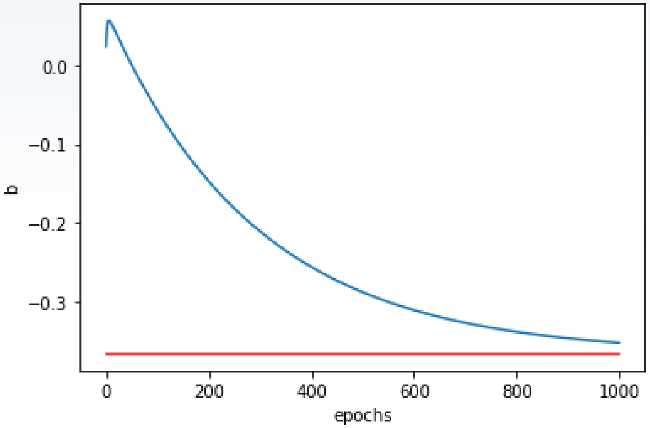
```
for i in range(1000):
  dw_ = dw(w,b)
  db_ = db(w,b)

  w = w - alpha * dw_
  b = b - alpha * db_

plot_reg_line(w,b)
```



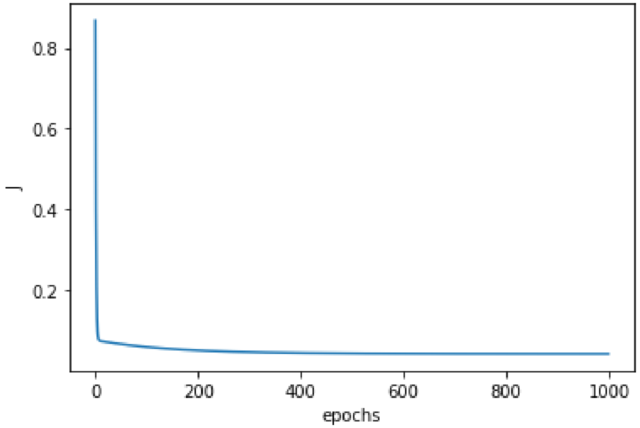
תהליך הלמידה

- פרמטרי המודל:



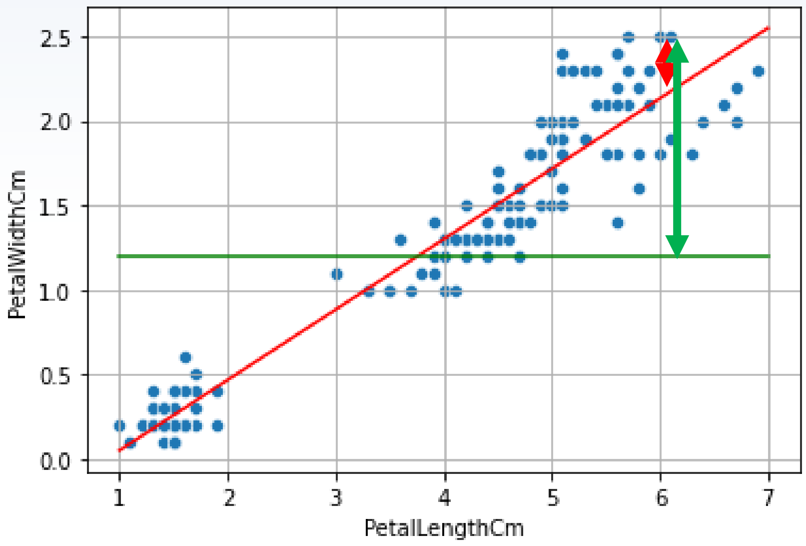


- פונקציית המחיר J



הערכת ביצועים ברגרסיה לינארית

הערכת ביצועים של רגרסיה לינארית
מדד השונות המוסברת R^2



$$SS_{tot} = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \bar{y}\right)^2$$

$$SS_{error} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$R^2 = 1 - \frac{SS_{error}}{SS_{tot}}$$

ערכים נעים (לרוב) בין 0-1

1: התאמה מלאה

0: אין התאמה

ערכים שליליים: חיזוי יותר גרוע מאשר מדד הממוצע

תרגיל 2 – רגרסיה עם GD

- ממשו את הקוד במצגת לפתרון רגרסיה לינארית עם GD (חיזוי petal width לפי petal length)
- שרטטו את גרפי ההתכנסות (w,b,J)
- השוו את קצב ההתכנסות עבור מספר ערכי alpha שונים
- מהו ערך alpha המקסימלי האפשרי?
- מה קורה אם alpha קטן מאוד?
- מהו מדד השונות המוסברת R^2

תרגיל 3

- נתונה רגרסיה לינארית עם מאפיין אחד ובסיס נתונים עם m דוגמאות.
- רשמו את הסוג (סקלר, וקטור, מטריצה) והממדים של המשתנים הבאים:

x

y

$\hat{y}$

w

b

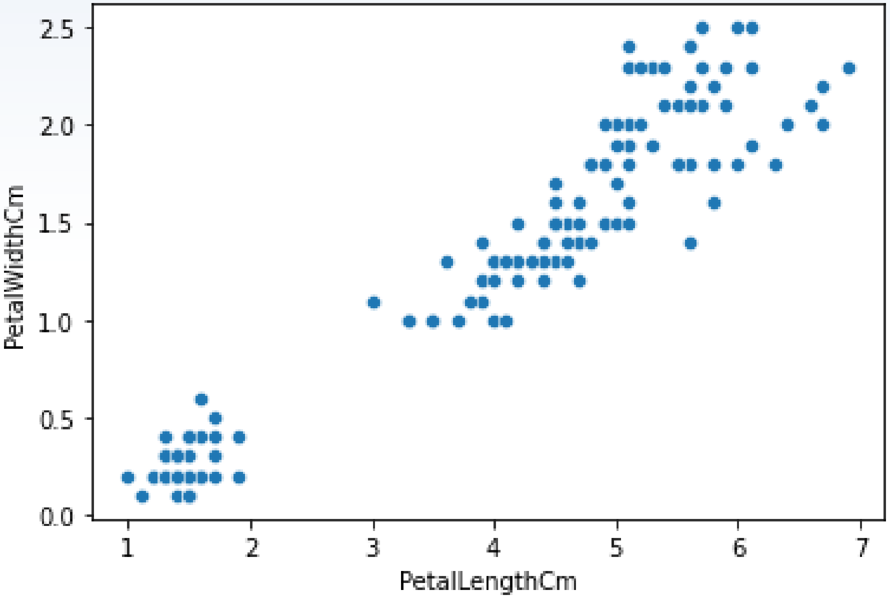
J

$\frac{\partial}{\partial w}J(w,b)$

$\frac{\partial}{\partial b}J(w,b)$

רגרסיה לינארית עם ספריית scikit-learn

דוגמא – חיזוי רוחב עלה של אירוס לפי אורכו



רגרסיה ליניארית עם SKLEARN

```
X=df[['PetalLengthCm']].to_numpy()
y=df['PetalWidthCm'].to_numpy()
```

```
from sklearn.metrics import r2_score
from sklearn.linear_model import LinearRegression
```

```
X1 = np.expand_dims(X,axis=1)
X1.shape
```

```
(150, 1)
```

```
reg = LinearRegression()
reg.fit(X1,y)
w_sk = reg.coef_
b_sk = reg.intercept_
print(w_sk,b_sk)
```

```
[0.41575542] -0.3630755213190291
```

חיזוי

```
# predict
reg.predict([[3]])
```

```
array([0.88274335])
```

הערכת ביצועים r^2

```
reg.score(X1,y)
```

```
0.9271098389904927
```

ויזואליזציה

```
sns.scatterplot(x=X, y=y)
r = np.array([[X.min()], [X.max()]])
r_pred = reg.predict(r)
plt.plot(r, r_pred, color='red')
plt.title("Scatterplot with Regression Line")
plt.xlabel("petal_length")
plt.ylabel("petal_width")
plt.show()
```