

ראסיה ליניארית - חלק ב'

טובת התצוי - מדד R^2

סקור שיקר חיוני, כי אחר נוסחת פיתוח השוני:

$$\sum (y_i - \bar{y})^2 = \underbrace{\sum (y_i - \hat{y}_i)^2}_{\text{השנוי (הלא מוסברת)}} + \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{השנוי המוסברת}}$$

ניתן להבין, כי מודל רגורסה סודי. תמו זווה שבו השנוי (הלא מוסברת) תהיה כמה שיותר קטנה והשנוי המוסברת תהיה כמה שיותר גדולה (כי זה באומר שהצלתו ארוסדיו סוג אחר השנוי ב- y ה- x ים שלנו).
אכן, מדד מקובל לטובת התצוי יהיה:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum e_i^2}{\frac{1}{n} \sum (y_i - \bar{y})^2}$$

אחוז השנוי המוסברת

סדק גדולה של $R^2 =$ משמעות מודל רגורסה טוב (כי הצלתו ארוסדיו את התצוי של המשתנה המוסבר).

לכן, נמוך של $R^2 =$ משמעות מודל רגורסה לא טוב.

$$R^2 = r^2(x, y) = \hat{\alpha}^2 \frac{\text{Var } x}{\text{Var } y}$$

מקדם

תצויות תרבות - מדד RMSE

(תצויות)

$$1. \quad MSE = \frac{1}{n-2} \sum e_i^2 \quad \text{אחוז השנוי (השאריות) - אחוז השנוי}$$

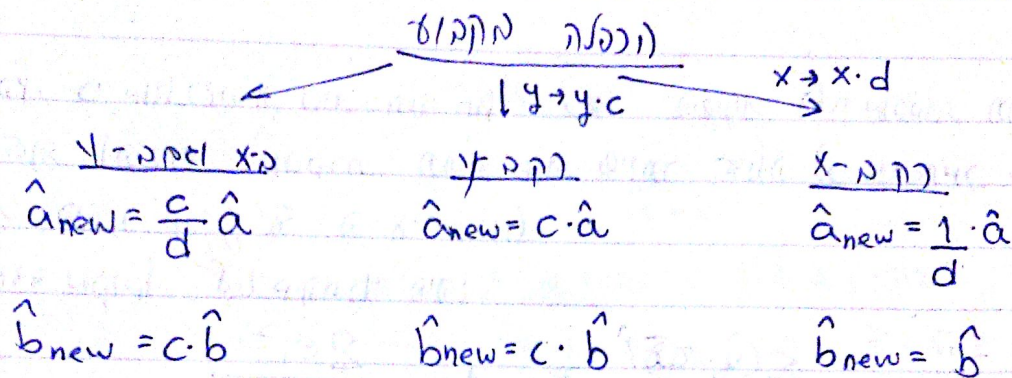
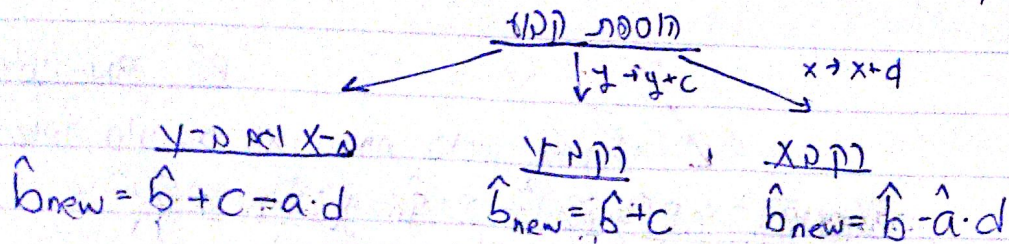
$$2. \quad RMSE = \sqrt{\frac{1}{n-2} \sum e_i^2} \quad \text{אחוז השנוי (השאריות) - אחוז השנוי}$$

3. תצויות תרבות - אחוז נקודות התצויות (השאריות) של המשתנה המוסבר.

$$ze_i = \frac{e_i}{RMSE} \leftarrow \text{אחוז השנוי}$$

היה גדול יחסית.

ע"י קנה מידה



הוכחות

1. נניח כי אני מחזים את הטנספורמציה:
 $x \rightarrow x+d$
 $y \rightarrow y+c$

אני והמטרה החדשה תהיה:

$$\hat{b}_{new} = \bar{y}_{new} - \hat{a}(\bar{x}_{new}) = (y+c) - \hat{a}(x+d)$$

המטרה החדשה:

$$\hat{a}_{new} = \frac{Sd_{y,new}}{Sd_{x,new}} \cdot r(x_{new}, y_{new})$$

$$= \frac{Sd_y}{Sd_x} \cdot r(x, y) = \hat{a} \rightarrow \text{אם המטרה החדשה}$$

הצגת הנתונים

$$z_{xi} = \frac{x_i - \bar{x}}{s_{Dx}}$$

הצגת הנתונים של x' ו-y'

$$\begin{cases} \sum z_x = 0 \\ \sum z_y = 0 \end{cases} \quad \text{ומתקיים:}$$

מתכונות של z_x ו- z_y הן זהות ל- x ו- y :

$$\frac{z_x \cdot r(x,y)}{\substack{\uparrow \\ \text{שה-} x' \text{ וה-} y' \\ \text{הן זהות}}} = \frac{\hat{y}}{\substack{\uparrow \\ \text{שה-} y' \\ \text{הן זהות}}}$$

מכיון שקולציה היא סימטרית ה- x ו- y נקבל כי r קו הריגורס של z_x על z_y יהיה:

$$z_y \cdot r(x,y) = \hat{z}_x$$

(כלומר מקבלים את אותו המשוואה).

תוצאות

1. (חשוב לדעת R שראוי לתאר הקדם. x ו- y נמצא נתיבם. y - תוצר במדויק).
כלי נמצא את אחוז השונות המוסברת. (R^2)

$$\hat{R}^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} \quad \text{או} \quad \text{summary(lm(y ~ x))}$$

2. מציא את האומדן לסטיית התקן של השאריות (RMSE).
= אינו מדויק הוא חייב? (הסטיית אומדן - שגיאת אומדן?)

$$\text{summary(lm(y ~ x))} \rightarrow \text{Residual standard error}$$

הפרדתם שנתן מאגריהם. לא זהות
כל אופן מציא תוצר.

2. נניח כי אנו מקבלים את התרגומים:

$$x \rightarrow x \cdot d$$

$$y \rightarrow y \cdot c$$

אני הוספתי גם הסימן:

$$\hat{a}_{new} = \frac{Sd_{y,new}}{Sd_{x,new}} \cdot r(x_{new}, y_{new})$$

$$= \frac{|c| Sd_y}{|d| Sd_x} \cdot r(x, y) \cdot \text{sign}(d \cdot c) = \frac{c}{d} \hat{a}$$

$$\hat{b}_{new} = \bar{y}_{new} - \hat{a}_{new} \cdot \bar{x}_{new} = c \cdot \bar{y} - \hat{a}_{new} \cdot \bar{x} \cdot d$$

$$= c \cdot \bar{y} - \frac{c}{d} \cdot \hat{a} \cdot \bar{x} \cdot d = c(\bar{y} - \hat{a} \bar{x})$$

בטו מקרה, צריך האקסצור (המחלק) בקלות מסופק;

1. דרגות:

$$x \rightarrow ax + b$$

$$y \rightarrow cy + d$$

2. דרגים בנוסחאות (המקדמים):

$$\hat{a}_{new} = \frac{Sd_{y,new}}{Sd_{x,new}} \cdot r(x_{new}, y_{new})$$

$$\hat{b}_{new} = \bar{y}_{new} - \hat{a}_{new} \cdot \bar{x}_{new}$$

3. לטוב - של מושגים מהכפלה - 1 ח מושג מ ממשותף להחכמים.

2. מה יקרה אם נחלק את התוצר ב-1000?

$$a_{\text{new}} = \frac{a}{1000}$$

$$b_{\text{new}} = \frac{b}{1000}$$

2. ספנינג נתונים איכותיים וזמן של סרטן המעי הגד של נשים הישראליות:

מספר נשים	שנה	
9.2	1984	ממוצע
5.02	17.25	סטיית תקן

$$\sum (\hat{y}_i - \bar{y})^2 = 1.56$$

מה סיכוי של מקרים המטות בין השנה לבין מספר הנשים התולדות?

$$r^2 = R^2 = 1 - \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{1.56}{(5.02)^2} = 0.9961$$

3. מורה בבית הספר תיכון, החליט לבדוק כיצד הציונים במבחן הבית המשותף והציונים במבחן ההיסטוריה (y) קשורים, והוא רצה אסוף נתונים מ-12 תלמידים וקובע כי:

$$\sum x_i = 1032 \quad \sum y_i = 1092$$

$$\sum x_i y_i = 75,560 \quad \sum x_i^2 = 2040 \quad \sum y_i^2 = 2000$$

א. מצא את ישר הרגרסיה:

$$\hat{a} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2} = \frac{\frac{75,560}{12} - \frac{1032}{12} \cdot \frac{1092}{12}}{\frac{2040}{12} - \left(\frac{1032}{12}\right)^2} = 0.211$$

$$\hat{b} = \bar{y} - \hat{a} \bar{x} = \frac{1092}{12} - 0.211 \cdot \frac{1032}{12} = 72.854$$

$$\rightarrow \boxed{\hat{y} = 72.854 + 0.211x}$$

ה. מצא את R^2

$$R^2 = r^2 = \frac{\hat{a}^2 \cdot \text{Var}x}{\text{Var}y} = 0.211 \cdot \frac{(\bar{x^2} - \bar{x}^2)}{(\bar{y^2} - (\bar{y})^2)}$$

$$= 0.211 \cdot \frac{-7.226}{-7.229} = 0.21$$

- ג. אלוו מורה, ג'לד, כ...
 1. המורה לבינה החליט להוריד לבסוף את הציון ג-12
 נקודות כי ג'לד סתמיזם הסתיקן.
 2. המורה להספורה החליט לתת לבסוף מוספת 10%
 לציון כי המבחן היה קשה מדי.

מה יהיו מקדמי הרגססיה?

$$x \rightarrow x - 12$$

$$y \rightarrow 1.1y$$

$$\rightarrow \hat{a}_{\text{en}} = \frac{1.1 \cdot s_{dy}}{s_{dx}} \cdot (+r_{(x,y)}) = 1.1 \cdot \hat{a}_{\text{en}} = 1.1 \cdot 0.211 = 0.232$$

↑
הסימן חיובי

$$\begin{aligned} \rightarrow \hat{b}_{\text{en}} &= \bar{y}_{\text{en}} - \hat{a} \cdot \bar{x}_{\text{en}} = 1.1 \cdot \bar{y} - 1.1 \hat{a} \cdot (\bar{x} - 12) \\ &= 1.1 \bar{y} - 1.1 \hat{a} \bar{x} + 3.2 \hat{a} \\ &= 1.1 \hat{b} + 3.2 \hat{a} = 80.816 \end{aligned}$$