

ביצועים של אלגוריתם - השוואה

תוכנית: הכוונה הקובעת - ציון ביצועים accuracy, recall, precision

* סוג אלגוריתם

* סוג ביטוי: חלק א': אלגוריתם שמתיר 1/0, חלק ב': איתני-ציון

* תמונה בעלת של הסכמה או אי הסכמה בין אלגוריתם

* עד כמה חלוקים

* האם הבדל בין אי הוא לשימוש?

* האם הסכמה בין אלגוריתם היא אקראית?

* האם יש קשר לזאת בין השימוש?

Contingency Table

למדידת יעילות, מסמכת הבדלים בין אלגוריתם

$$n_{11} + n_{10} + n_{01} + n_{00} = M$$

$$Accuracy_1 = \frac{n_{11} + n_{10}}{M}$$

$$Accuracy_2 = \frac{n_{11} + n_{01}}{M}$$

* ערכים n_{10} , n_{01} מראים אי הסכמה

ציון הסכמה

$M=200$

$$Acc_1 = \frac{150+25}{200} = 87.5\%$$

$$Acc_2 = \frac{150+25}{200} = 82.5\%$$

הפרש של 5% > > ביצועים

Classifier 1 (C_1)

	Correct, $\hat{Y}_1 = Y$	Incorrect, $\hat{Y}_1 \neq Y$
$\hat{Y}_2 = Y$	Both correct $\hat{Y}_1 = Y = \hat{Y}_2$ $\hat{Y}_1 = \hat{Y}_2$ n_{11} שניהם צודקים	C_1 incorrect, C_2 correct $\hat{Y}_1 \neq Y = \hat{Y}_2$ n_{01}
$\hat{Y}_2 \neq Y$	C_1 correct, C_2 incorrect $\hat{Y}_1 = Y \neq \hat{Y}_2$ n_{10}	Both incorrect $\hat{Y}_1 \neq Y \neq \hat{Y}_2$ $\hat{Y}_1 = \hat{Y}_2$ n_{00}

Classifier 1

	Correct	Incorrect
Correct	150	15
Incorrect	25	10

הפרש של 5%

Disagreement Measure

$$Dis = \frac{25 + 15}{200} = \frac{40}{200} = 0.2$$

עצם כמה תלוקים בצעות?

$$Dis = \frac{n_{10} + n_{01}}{M}$$

McNemar's Test

האם הבדל בין $n_{10} \leftarrow n_{01}$ יכול להיות במקרה?

Null hypothesis $H_0: n_{10} = n_{01}$ - השאנה שבצעות זהה

$H_1: n_{10} \neq n_{01}$

p-value ← test statistic $\frac{(n_{10} - n_{01})^2}{n_{10} + n_{01}}$

α רמת החשיבות

צותים $H_0 \rightarrow \alpha < p$

נשארים עם $H_0 \rightarrow \alpha \geq p$

עכשיו צינאה לפסגות

$$\frac{(25 - 15)^2}{25 + 15} = \frac{100}{40} = 2.5 \rightarrow p \approx 0.114$$

ערכים של $\alpha = 0.01$ ו- 0.05 $\Leftarrow \alpha < p \Leftarrow$ לא ניתן לעזרת H_0
 \Leftarrow לא ניתן לעזרת H_0 שאיננו
 בעל הבדל משמעותי בצעות



עצם כמה שונים (האם משמעותי)?

↓ עצם כמה מסכמים?

Cohen's Kappa Coefficient

הסתברות להסכמה
 בין שני מסווגים

$$p_o = \frac{n_{11} + n_{00}}{M}$$

observed

ציון של שני מסווגים

$$Accuracy_1 = \frac{n_{11} + n_{10}}{M}$$

$$Accuracy_2 = \frac{n_{11} + n_{01}}{M}$$

$$p_e = \underbrace{Accuracy_1 \cdot Accuracy_2}_{\text{both correct}} + \underbrace{(1 - Accuracy_1)(1 - Accuracy_2)}_{\text{both incorrect}}$$

↓
 expected

הכנתה של
 א לטו

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

- $\kappa = 1$ הסכמה מלאה
- $\kappa = 0$ הסכמה אקראית
- $\kappa < 0$ אי הסכמה שיטתית

$$p_o = \frac{150 + 10}{200} = 0.80$$

$$\text{Accuracy}_1 = \frac{150 + 25}{200} = 0.875, \quad \text{Accuracy}_2 = \frac{150 + 15}{200} = 0.825$$

$$p_e = 0.875 \times 0.825 + 0.125 \times 0.175 = 0.7219 + 0.0219 = 0.7438$$

$$\kappa = \frac{0.80 - 0.7438}{1 - 0.7438} = \frac{0.0562}{0.2562} \approx 0.219 \leftarrow \text{קצב מעט אקראי}$$

הערה: א לא קשור עב'צוים של ס מסוג, אלא ס הסכמה ביניים

Q-Statistic (Yule's Q)

דקטור בין שני מסווגים עם כוונות

Range: $Q \in [-1, 1]$.

תוצאות
אם
שני
תוצאות
הן
אלו
הן

- $Q = 1$: the classifiers are perfectly positively associated (whenever one is correct, so is the other).
- $0 < Q < 1$: the classifiers are positively correlated, meaning they tend to recognize the same instances correctly and incorrectly.
- $Q = 0$: the classifiers are independent. \rightarrow בלתי תלויים
- $-1 < Q < 0$: the classifiers are negatively correlated, meaning they tend to make errors on different instances.
- $Q = -1$: perfect negative association (one is correct if and only if the other is incorrect).

$$Q = \frac{n_{11} n_{00} - n_{10} n_{01}}{n_{11} n_{00} + n_{10} n_{01}}$$

$$Q = \frac{150 \times 10 - 25 \times 15}{150 \times 10 + 25 \times 15} = \frac{1500 - 375}{1500 + 375} = \frac{1125}{1875} = 0.6 \leftarrow \text{קטור חיובי}$$

השוואה בין מסווגים בעלי אורך הסתברות

$$p_k = f_{\theta}(x) = \Pr(\hat{y}=1)$$

K מספרם של מסווגי

* אם כמה מסווגי תוצאה איתנו

הסתברות \rightarrow תוצאה

* האם הבדל יורם עדיף במקרה?

* האם נופלים על אותם נטות?

Brier Score

שיטה: MSE בין הסתברות לעדתי y_i

$$0 < BS < 1$$

$$BS = \frac{1}{M} \sum_{i=1}^M (p_i - y_i)^2$$

נושאם $p_i = y_i$

$$p_i = 1/2 \Rightarrow BS = 1/4$$

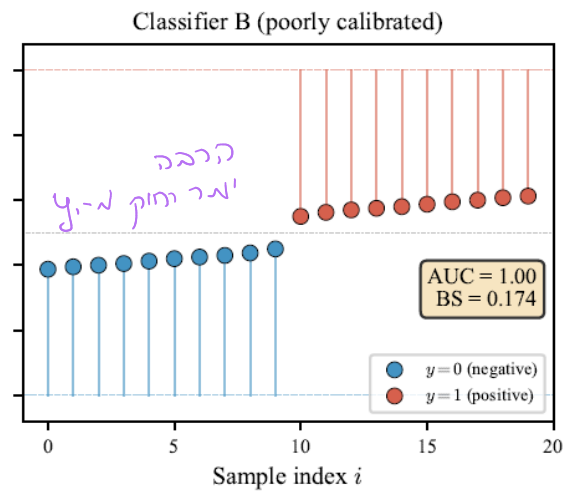
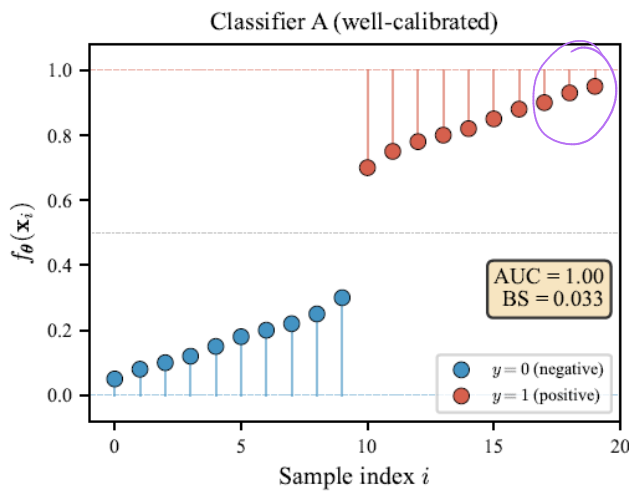
ערך איתנו

y_i	\hat{y}_i	$(p_i - y_i)^2$
1	0.9	0.01
0	0.2	0.04
1	0.7	0.09
0	0.1	0.01

$$BS = \frac{0.01 + 0.04 + 0.09 + 0.01}{4} = 0.0375$$

קדום 0-8
כה טוב!

נושא מסווגים איתנו



Paired Scoring Rule Comparison

הסתברות p ו- y

הסתברות $1-p$ ו- $1-y$

הסתברות p ו- y (ערכים אחרים)

$$S_{BS}(p, y) = (p - y)^2$$

$$S_{LL}(p, y) = -[y \log p + (1-y) \log(1-p)]$$

$$S_{k,i}$$

נושא

Score

$p_{k,i}$ ו- y_i

i - מספר שורה בבסיס נתונים

k - מספר ה- S מסווג

$$\delta_i = S_{1,i} - S_{2,i}$$

$$\bar{\delta} = \frac{1}{M} \sum_{i=1}^M \delta_i = \bar{S}_1 - \bar{S}_2$$

$$\delta > 0 \quad \text{מ' 2}$$

$$\delta < 0 \quad \text{מ' 1}$$

$$\bar{S}_k = \frac{1}{M} \sum_{i=1}^M S_{k,i}$$

y_i	$p_{1,i}$	$S_{BS,1}$	$p_{2,i}$	$S_{BS,2}$	δ_i
1	0.90	0.010	0.75	0.063	-0.053
0	0.20	0.040	0.10	0.010	+0.030
1	0.70	0.090	0.85	0.023	+0.068
0	0.30	0.090	0.40	0.160	-0.070
1	0.60	0.160	0.80	0.040	+0.120
0	0.15	0.023	0.25	0.063	-0.040
$\bar{S}_1 = 0.069$		$\bar{S}_2 = 0.060$		$\bar{\delta} = +0.009$	

הסתברות קטנה של 2

עצמאי

Paired t-test on Score Differences

אם $\bar{\delta}$ הוא משמעותי ל-0

$$\alpha \geq p\text{-value} \leftarrow \frac{\bar{\delta}}{s_{\delta}/\sqrt{M}} \leftarrow \text{מקובל}$$

0.01

$$p \approx 0.77 > \alpha \Rightarrow \text{הבדל לא משמעותי}$$

$< \alpha$

Wilcoxon Signed-Rank Test on Score Differences

δ_i	$ \delta_i $	Rank R_i	Sign
+0.030	0.030	1	+
-0.040	0.040	2	-
-0.053	0.053	3	-
+0.068	0.068	4	+
-0.070	0.070	5	-
+0.120	0.120	6	+

$\delta_i \geq 0$ & חצי יותר קטן
 & חצי יותר יאר
 לזוסי ס'אני & יב'אס
 סי'אני & יב'אס

$p > \alpha$ בקרה
 $p < \alpha \rightarrow$ ח'יון שונה
 אפס על בקרה