Lecture-4 (Performance Measures)

Friday, 11 August 2023 4:30 PM

Actual \ Predicted	+	−
+	TP	FN
−	FP	TN

$$\text{Precision } (P) = \frac{TP}{TP+FP}$$

$$\text{Recall } (R) = \frac{TP}{TP+FN}$$

- $0 < P < 1, \quad 0 < R < 1$
- Precision is undefined : $TP + FP = 0$
  → Classifier predicts everything as −
- Recall is undefined : $TP + FN = 0$
  → Samples are from − class only
- $P = 1 \Rightarrow \frac{TP}{TP+FP} = 1 \Rightarrow TP = TP + FP \Rightarrow FP = 0$
- $R = 1 \Rightarrow FN = 0$
- ~~$\frac{P+R}{2}$~~ $\quad F1\text{-score} = \frac{2PR}{P+R} = \frac{TP}{TP + \frac{1}{2}(FP+FN)}$

Multi-Class Classification - Performance Measures

confusion matrix

Predicted \ Actual	Urgent	Normal	Spam
Urgent	8	10	1
Normal	5	60	50

1 - vs - rest

Spam	3	30	200

Urgent

Predicted

	Urgent	Not Urgent
Urgent	8	11
Not Urgent	8	340

$$P_{Urgent} = \frac{8}{8+11} = \frac{8}{17}$$

$$R_{Urgent} = \frac{8}{8+8} = \frac{8}{16}$$

Predicted

	Spam	Not Spam
Spam	200	33
Not Spam	51	83

$$P_{Spam} = \frac{200}{233} \approx 0.91$$

$$R_{Spam} = \frac{200}{251} \approx 0.88$$

Predicted

	Normal	Not Normal
Normal	60	55
Not Normal	40	212

$$P_{Normal} = \frac{60}{115} = 0.52$$

$$R_{Normal} = \frac{60}{100} = 0.6$$

macro score

$$P_{macro} = \frac{P_{Urgent} + P_{Normal} + P_{Spam}}{3} = \frac{\sum_{i=1}^{} P_i}{N}$$

Pooled

Pred \ Actual	+	−
+	268	99
−	99	635

$$P_{micro} = \frac{268}{268+99}$$

$FP = FN$

FP GT Pred

Normal	Spam

Spam

$$F1_{micro} = \frac{2 P_{micro} R_{micro}}{P_{micro} + R_{micro}}$$

$$F1_{macro} = \frac{\sum_{i=1}^{N_c} F_i}{N_c}$$

scikit-learn

ALT

$$\frac{2 P_{macro} R_{macro}}{P_{macro} + R_{macro}}$$

$$F1\text{-score} = \frac{TP}{TP + \frac{1}{2}(FP+FN)}$$

$$F1\text{-score}_{micro} = \frac{\sum_{i=1}^{N_c} TP_i}{}$$

sample-centric

$$\sum_{i=1}^{N} TP_i + \frac{1}{2}\left(\sum FP_i + \sum FN_i\right)$$

	TP	FP	Prec
A	1	1	0.5
B	1	1	0.5
C	90	1	≈ 1
D	1	1	0.5

$Prec_{macro} \approx 0.62$

$Prec_{micro} = \frac{93}{[illegible]}$

$\frac{93+4}{} \approx 0.95$

---

* $n/2$ Classifier : Random Guessing

$-$ $n/2$ $P(Pred = +) = \frac{1}{2}$

$P(Pred = -) = \frac{1}{2}$

Accuracy = $\frac{\#\text{ of correct predictions}}{n}$ → R.V → $\frac{(n/2)}{n} = \frac{1}{2}$

+ + + + + − − − − −

✓ + + + + + + + + + +

✓ + + ... + −

$E[\#\text{ of correct predictions}]$

↓

$0 \ldots n \quad [0, n]$

(n) (p)

How many times do you get k successes?

Binomial distribution

$E[\cdot] = n \cdot p$

$= n\left(\frac{1}{2}\right)$