

## Applications du cours

### Comparer des modèles

Deux modèles prédictifs ont été choisis et appliqués sur une population test. Les résultats sont donnés dans le tableau suivant :

	Indiv. X'	Ref.	Hyp. 1	Hyp. 2
2 a	1	+	+	+
2 d	2	+	+	+
3 d	3	+	+	-
c	4	-	+	-
u a	5	+	+	+
c a	6	+	+	+
6 a	7	+	+	-
c	8	-	+	-
d	9	-	-	-
d	10	-	-	-

- Quel type d'erreur peut-on évaluer ?
- Calculer les tables de contingence. *→ calcul de  $I_c$ , pas utile car donnée*
- Quel est le taux de mauvaise classification pour chacun des systèmes ?
- Donnez la précision et le rappel des modèles 1 (Hyp1) et 2 (Hyp2).  
Précisez également le détail pour chaque classe.
- Quel est selon vous le meilleur système ? Justifiez. *( si  $P$  et  $R$  de A > à celui de B ; A est probablement le meilleur )*
- Avez-vous considéré la significativité de vos résultats ?

Aide :

$$CER = \frac{\# \text{ instances mal classées}}{\# \text{ instances classées}}$$

$$CER \pm 1.96 \sqrt{\frac{CER \cdot (1 - CER)}{\# \text{ instances classées}}}$$

$$\text{précision}_i = \frac{\# \text{ instances correctement classées } i}{\# \text{ instances classées } i}$$

$$\text{rappel}_i = \frac{\# \text{ instances correctement classées } i}{\# \text{ instances réellement } i}$$

$$\text{précision} = \frac{\sum_i \text{précision}_i}{\text{nombre de classes}}$$

$$\text{rappel} = \frac{\sum_i \text{rappel}_i}{\text{nombre de classes}} \Rightarrow \text{pertinence}$$

$$f_{\text{mesure}} = \frac{(1 + \beta^2) \text{rappel} * \text{précision}}{\beta^2 (\text{rappel} + \text{précision})}$$

*nombre de fois où j'ai eu raison*  
*> nombre de fois que j'ai dit "+" ou "-" ou autre*

*-  $\Sigma$  peut avoir 90% de bonne réponses  
 - et trouver 10% de réponses possibles*

# Exo 2 Louis

ANN 11-03-2020

- a.) - recherche prédictive (vrai)  
- erreur en généralité (c'est généralisation, car on fait tout sur le corpus Test)

b.)

		Hyp 1	
		+	-
Act	Hyp 1	a 6	b 0
	Ref	c 2	d 2

ref: 6  
- 4 -

		Hyp 2	
		+	-
Ref	Hyp 2	e 4	f 2
	Act	g 0	h 4

→ On a 10 données: somme des cases = 10

c.)  $CER_1 = \frac{\text{nombre d'erreurs ce qui est dehors de diag}}{10} = \frac{2}{10}$  ( $CER_2 = \frac{2}{10}$ )

d.) Ici on a taux d'erreur pas besoin car le taux très peu de données.

On va le faire quand m.

$$20\% \pm 196 \times \sqrt{\frac{0.2 \times 0.8}{10}}$$

Précision: pour le  $E_1$ .

je raisonne  
j'ai raison  
qd je dis  
quelque chose

$$P_1 = \frac{b(+)}{b(+)+c(+)} = \frac{6(+)}{8(+)} = 75\%$$

nombre de (+)

$$P_1 = \frac{2}{2} = 100\%$$

$$P_2 = \frac{\sum P}{n} = 97,5\%$$

mbre de fois ou qui est  
représenté  
mbre de fois j'ai de

Rappel  $r_1 = \frac{b}{b+d} = 100\%$

ou avec  
Ref

$$r_2 = \frac{2}{4} = 50\%$$

combien de

taux d'erreur sont les m

Dans l'adéquation de ce qui est privilégié précision # Rappel