

Apprentissage numérique – M1 Informatique

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
1	+	+	+
2	+	+	-
3	+	+	-
4	-	+	-
5	-	-	-

- Quel type d'erreur peut-on évaluer ?
- Calculer les tables de contingence.
- 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.
- Avez-vous considéré la significativité de vos résultats ?

Aide :
$$E' \pm 1.96 \sqrt{\frac{E'(1 - E')}{i}}$$

Evaluation

On souhaite associer chaque individu d'une population donnée à l'une des classes suivantes : A, B, C ou D. Soit X un ensemble de 14 individus que l'on observe pour l'apprentissage et X' un ensemble de 6 individus que l'on observe pour le test. Deux modèles sont appris, le premier modèle prédit les hypothèses 1 et le second les hypothèses 2. Les résultats sont donnés dans les tableaux ci-après

Indiv. X	Réf.	Hyp. 1	Hyp. 2
1	A	A	A
2	A	C	A
3	B	B	C
4	C	C	B
5	C	C	C
6	A	A	A
7	D	D	C
8	A	A	A
9	A	A	B
10	C	C	C
11	C	C	C
12	C	C	C
13	C	C	B
14	A	C	A

Indiv. X'	Réf.	Hyp. 1	Hyp. 2
15	C	A	C
16	A	A	A
17	C	C	B
18	A	A	A
19	B	C	B
20	D	D	C

2. Quel est le type de tâche ?
3. Ecrire la table de contingence « Tab_Ep1 » qui permet de calculer l'erreur de prédiction du modèle 1.
4. Calculer sur *Tab_Ep1* :
 - a. Taux d'erreur global du système 1
 - b. Mesure Kappa

Aide :

$$kappa = \frac{P_{accord} - P_{hasard}}{1 - P_{hasard}}$$

Avec

$$P_{hasard} = \frac{\sum_{k \in Y} \#ligne_k \cdot \#colonne_k}{N^2}$$

- c. Précision
 - d. Rappel
 - e. F-mesure
5. Quelles remarques faites-vous sur le pouvoir de prédiction du modèle 1 ?
6. Considérons la tâche du domaine de la recherche d'information. Quel est le modèle le plus performant côté généralisation?