

19 Septembre 2025

Classification.

1-) KNN et les métriques de classification

2-) Pratique KNN

- calcul KNN.

- calcul métriques.

.

3 - TP: Début EDA.

classification knn

$$\hat{y} = f(x_i)$$

$$y \in \{l_1, l_2, \dots, l_m\}$$

cible

labels de donne

x_1	x_2	x_n	y
		.	l_1
		..	l_2
			:

$$d(e_{\text{Rech}}, e_i) \rightarrow \text{labels}$$

les plus proches

du label de e_{Rech}

1) définir d : distance.

2) Calculer TOUTES les distances.

3) Trier par ordre croissant de distance

4) Garder les K distances les plus courtes

Vote
Egalité

5) Voter majoritaire $y_{\text{Rech}} \rightarrow$ Pmajoritaire / Egalité / Pondérée

- Recommandation : - Pas de phase de training.
- Coût de calcul élevé car on doit calculer toutes les distances.

Cahier

031-manip-KNN

- 1) Modèle KNN → Mesures Métriques.
- 2) Modèle Logistic → Mesures Métriques
Régression
- 3) Déterminer la valeur de k Adequate.

Regression

$$\frac{\text{RMSE}}{\text{R-squared}}$$

$$\begin{aligned}y_{\text{Test}} &= 102 \\y_{\text{Predict}} &= 104.5\end{aligned}$$

$$\frac{|y_{\text{predict}} - y_{\text{Test}}|}{y_{\text{real}}} = \frac{|104.5 - 102|}{102} = \underbrace{2.5}_{\sim}$$

Classification

$$y \in \{\text{rouge, vert, blanc}\}$$

$$y_{\text{Test}} (\text{real}): \underline{\text{vert}}$$

1 erreur

$$\begin{array}{ll}y_{\text{Predict}} : & \underline{\text{blanc}} \\ & \text{vert}\end{array}$$

0 erreur

100 erreurs → success

Erreurs de type I (faux positif)
Erreurs de type II (faux négatif).

Matrice de confusion.

↳ cahier 832

Test d' hypothèse

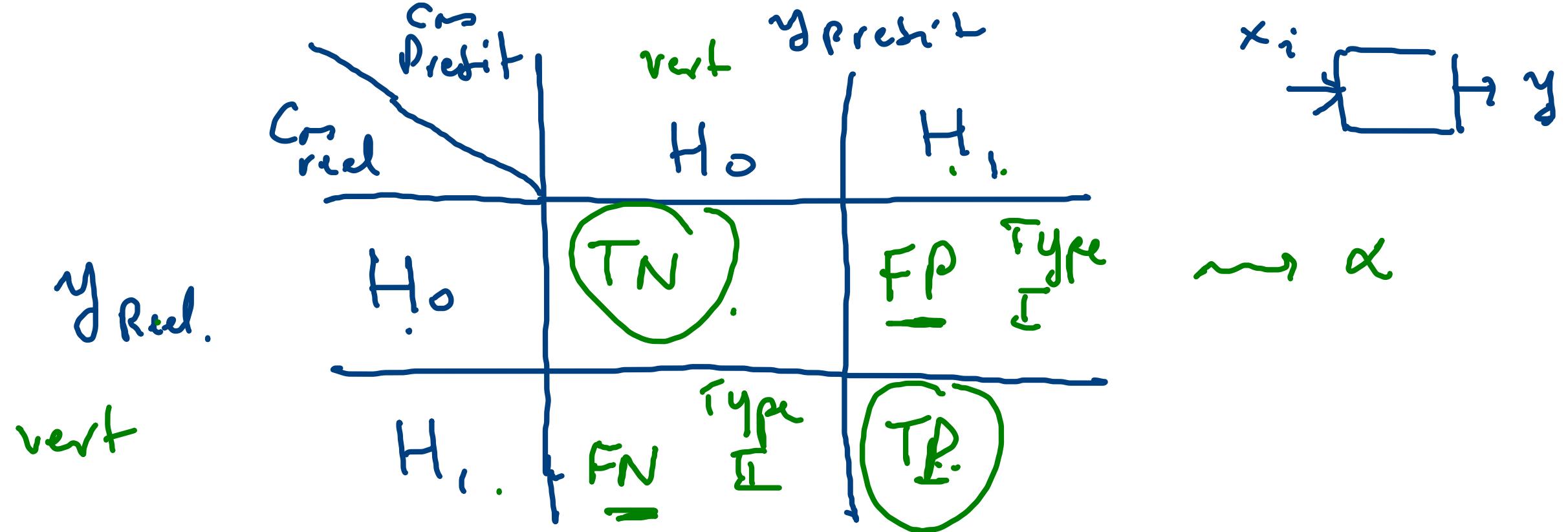
Négatif H_0 : médicament
n'a pas d'effet

Positif H_1 : médicament a
une effet

statistique
Échantillon.

Conclusion:

[accept H_0
refute H_0]



True Negative (TN)

True positive (TP)

False positive (FP)

False negative (FN)

1000

	P			
R	H ₀	H ₁		
H ₀	996	3		
H ₁	1	5		

Accuracy: $\frac{T_N + T_P}{\text{Total}} = \frac{991 + 5}{1000} = \underline{\underline{99.6\%}}$

Precision: $\frac{T_P}{T_P + F_P} = \frac{5}{5 + 3} = \frac{5}{8} = \underline{\underline{62.5\%}}$

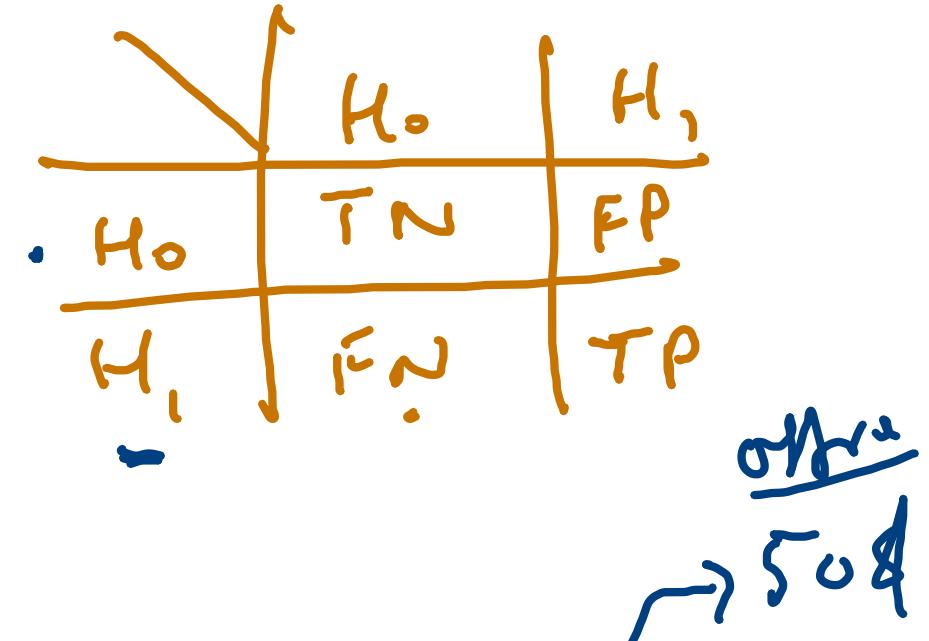
Sensitivity: $\frac{T_P}{T_P + F_N} = \frac{5}{5 + 1} = \frac{5}{6} = \underline{\underline{83.3\%}}$

Score F1 = $\frac{2 * \text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} = \underline{\underline{71.4\%}}$

Entreprise

H_0 : pas de fraude C.C.

H_1 : fraude



Tr: Reelle \rightarrow Modèle : H_1 fraude. \underline{FP} client

Tr: fraude.

↳ 1 M\$

Modèle : H_0

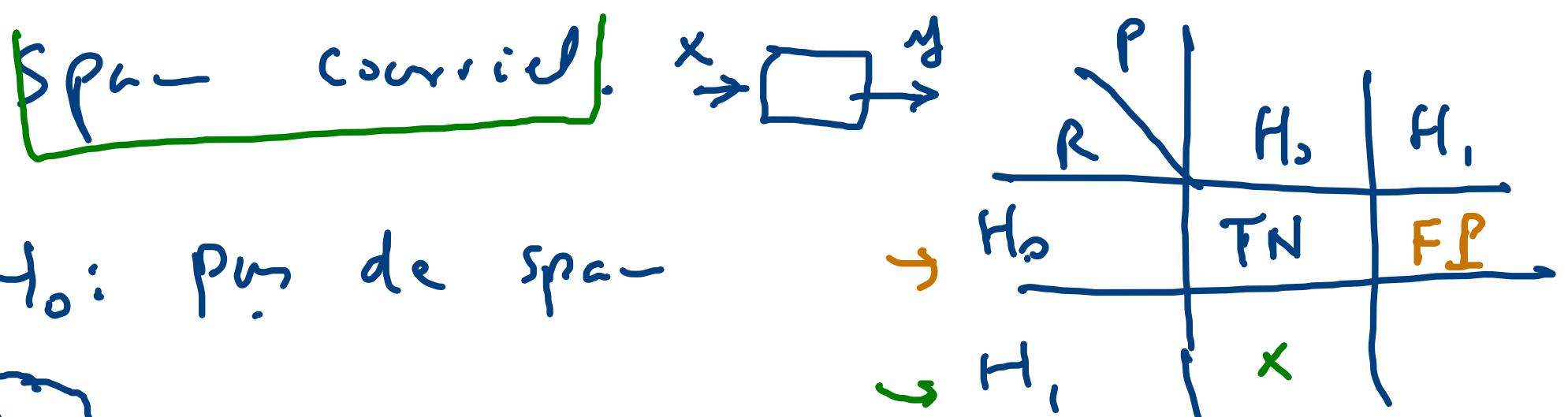
(FN)

Entreprise.

Min FP

M. Min FN

$$\frac{\text{Legibilité}}{\frac{TP}{TP+FN}} \downarrow 0$$



Confid. H_0 (pas de spm) $\rightarrow H_1$: spam

Loto Quebec $\xrightarrow{\text{long}}$ F.P.

Début

Confid. H_1 $\rightarrow H_0$

Prise $\xrightarrow{\text{long}}$ F_N

(spam)

Min FP

$P_{\text{recu}} = \frac{TP}{TP + FP}$

Temps

Test Hypothesis.

H_0 absence.

H_1 pas d'absence.

Procuren Couronne.

Avocat

H_0 : innocent

H_0 : Coupable

H_1 : pas innocent

H_1 : non coupable.

Personne,

Sante Canada

N H_0 : bon

P H_1 : pas bon

Rejet

Fabricant Medicant

N H_0 : pas bon.

P H_1 : bon

Prise en
stat d'opér.

Medicament

- Mehrige Technis.
 - Mehrige Business
-
- Qualität der
mehrige