

19 Septembre 2025

Classification.

1-) KNN et les métriques de classification

2-) Pratique KNN

- calculer KNN.

- calculer métriques.

3- TP: Debut EDA.

classification knn

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

→ prediction.

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

labels de classe

$d(e_{\text{Rech}}, e_i) \rightarrow$ labels
les plus proches
du label de e_{Rech}

The diagram shows a query point e_{Rech} with three green arrows pointing to a table. The table has columns for features x_1, x_2, \dots, x_n and a target label y . The first column is labeled x_1 , the second x_2 , and the last x_n . The last column is labeled y . The table contains several rows, with the first row having a dot in the x_2 column and the last row having a dot in the x_n column. The label y in the last column is shown as a list of labels l_1, l_2, \dots, l_m .

x_1	x_2	\dots	x_n	y
	.			l_1
				l_2
				l_3
				\vdots
				l_m

1) définir d : distance.

2) Calculer TOUTES les distances.

3) Trier par ordre croissant de distance.

4) Garder les K distances les plus faibles.

5) Vote majoritaire $y_{\text{Rech}} \rightarrow$ majoritaire

Vote
Égalité
Pondéré

Re-critique : - Pas de phase de training.

- Coût de calcul élevé car on doit calculer toutes les distances.

Calcul

031_mai_p_knn

1) Modèle KNN → Métriques

2) Modèle Logistic Regression → Métriques

3) Déterminer la valeur de k Adequate.

Regression

RMSE
R-squared

$$y_{\text{Test}} = 102$$
$$y_{\text{Predict}} = 104.5$$

$$| \underbrace{y_{\text{Predict}}}_{y_{\text{Predict}}} - \underbrace{y_{\text{Test}}}_{y_{\text{devel}}} | = | 104.5 - 102 | = 2.5$$

Classification

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

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

1 erreur

$y_{\text{Predict}} : \underline{\text{blanc}}$
vert

0 erreur

100 erreurs ——— erreurs
 ~ succès

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

Matrice de confusion.

↳ cahier 032

Test d'hypothèse

Négatif H_0 : médicament
nulle n'a pas d'effet

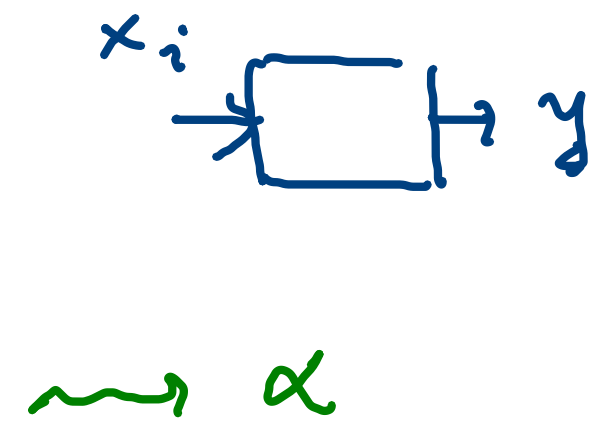
Positif H_1 : médicament a
alternation un effet

Statistique

Éducation.

Conclusion.
[accept H_0
rejette H_0]

		vert $y_{\text{predicted}}$	
		H_0	H_1
y_{Real}	H_0	\textcircled{TN}	\underline{FP} Type I
	H_1	\underline{FN} Type II	\textcircled{TP}



- True Negative (TN)
- True positive (TP)
- False positive (FP)
- False Negative (FN)

1000

R \ P	H ₀	H ₁
H ₀	991	3
H ₁	1	5

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

Precision: $\frac{TP}{TP + FP} = \frac{5}{5 + 3} = \frac{5}{8} = \underline{62.5\%}$

Sensitivity: $\frac{TP}{TP + FN} = \frac{5}{5 + 1} = \frac{5}{6} = \underline{83.3\%}$

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

Entreprise
 H_0 : pas de
 fraude C.C

H_1 : fraude

	H_0	H_1
H_0	TN	FP
H_1	FN	TP

offre
 $\rightarrow 508$

Tr: Reelle \rightarrow Mobile: H_1
 fraude. FP Client

Tr: fraude.

\hookrightarrow 1M€

Mobile: H_0

FN

Entreprise.

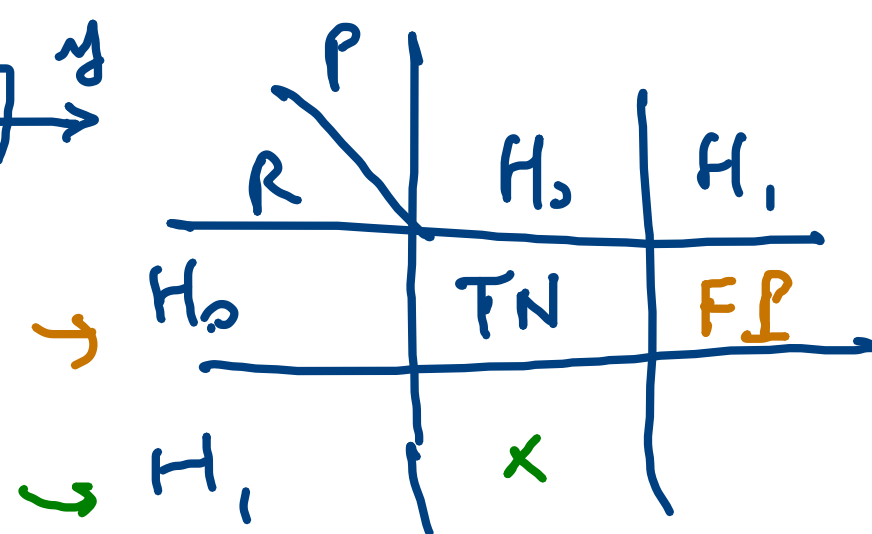
Min FP
Min FN

Sensibilité
 $= \frac{TP}{TP + FN}$
 \downarrow
 0

Spam - Courriel. $x \rightarrow \boxed{} \rightarrow y$

② H_0 : pas de spam

① H_1 : spam
 $x_1 \dots x_n$



Courriel H_0 (pas de spam) $\rightarrow \overline{H_1}$: spam Déduit
Loto Quebec 10M\$ FP

Courriel H_1 $\rightarrow H_0$: pas spam Temps
Prince 7M\$ (spam) FN
Min FP $= \frac{TP}{TP + FP}$
 \downarrow

Test Hypothese.

H_0 absence.

H_1 pas d'absence.

Procureur Couronne.

H_0 : innocent

H_1 : pas innocent

Avocat

H_0 : coupable

H_1 : non coupable.

Personne,

Sante Canada

(N) H_0 : bon

(P) H_1 : pas bon

✓
Rejet

Fabricant Medicament

(N) H_0 : pas bon.

(P) H_1 : bon

(Pharm.
et d-)

Medicament

- Métrique Technique.

- Métrique Business



Qualité du
modèle