

# Concepts mathématiques – Estimation de probabilité de COVID-19 à partir de symptômes

ANDRIATSIFERANA No Kanto Lorida

## 1. Modélisation du problème

On souhaite estimer la probabilité qu'un patient soit positif au COVID-19, étant donné les symptômes qu'il présente.

C'est un problème de classification binaire :

- Classe 1 : **COVID Positif** ( $y = 1$ )
- Classe 0 : **COVID Négatif** ( $y = 0$ )

Chaque symptôme est une variable binaire (présent ou absent), représentée dans un vecteur  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ .

## 2. Hypothèse : modèle de Naive Bayes

Nous utilisons le théorème de Bayes :

$$P(y = 1 | \mathbf{x}) = \frac{P(\mathbf{x} | y = 1) \cdot P(y = 1)}{P(\mathbf{x})}$$

Puisque  $P(\mathbf{x})$  est commun à toutes les classes, on peut comparer uniquement les numérateurs :

$$P(y = 1 | \mathbf{x}) \propto P(y = 1) \cdot \prod_{i=1}^n P(x_i | y = 1)$$

$$P(y = 0 | \mathbf{x}) \propto P(y = 0) \cdot \prod_{i=1}^n P(x_i | y = 0)$$

## 3. Estimation des probabilités conditionnelles

À partir du jeu de données :

- $P(y = 1)$  : proportion de cas positifs
- $P(x_i = 1 | y = 1)$  : fréquence du symptôme  $i$  chez les cas positifs
- $P(x_i = 0 | y = 1) = 1 - P(x_i = 1 | y = 1)$

Pour éviter les divisions par zéro, on applique un lissage de Laplace :

$$P(x_i = 1 | y) = \frac{n_i + 1}{n + 2}$$

## 4. Calcul pratique : log-vraisemblance

Pour éviter les sous-flux numériques, on calcule en logarithmes :

$$\log P(y = 1 \mid \mathbf{x}) = \log P(y = 1) + \sum_{i=1}^n \log P(x_i \mid y = 1)$$

$$\log P(y = 0 \mid \mathbf{x}) = \log P(y = 0) + \sum_{i=1}^n \log P(x_i \mid y = 0)$$

Ensuite, on applique la normalisation log-softmax :

$$P(y = 1 \mid \mathbf{x}) = \frac{e^{\log P_1}}{e^{\log P_1} + e^{\log P_0}}$$

## 5. Interprétation des résultats

- Plus un utilisateur coche de symptômes fréquents chez les cas positifs, plus la probabilité estimée augmente.
- Le score final est une probabilité entre 0 % et 100 %, facile à interpréter.
- On applique un seuil simple (ex. 50%) pour alerter ou rassurer l'utilisateur.

## 6. Exemple concret

Symptômes cochés : Toux sèche, Fièvre, Perte de goût.

—  $P(\text{Toux sèche} \mid y = 1) = 0,7$

—  $P(\text{Fièvre} \mid y = 1) = 0,8$

—  $P(\text{Perte de goût} \mid y = 1) = 0,75$

Calcul :

$$\log P(y = 1) + \log(0,7) + \log(0,8) + \log(0,75)$$

Et de même pour  $y = 0$ , puis normalisation.