

EPISEN – ING3. SI

Machine Learning



Abdallah EL HIDALI

Tech Lead Sita For Aircraft
abdallah.el-hidali@sit.aero

EPISEN

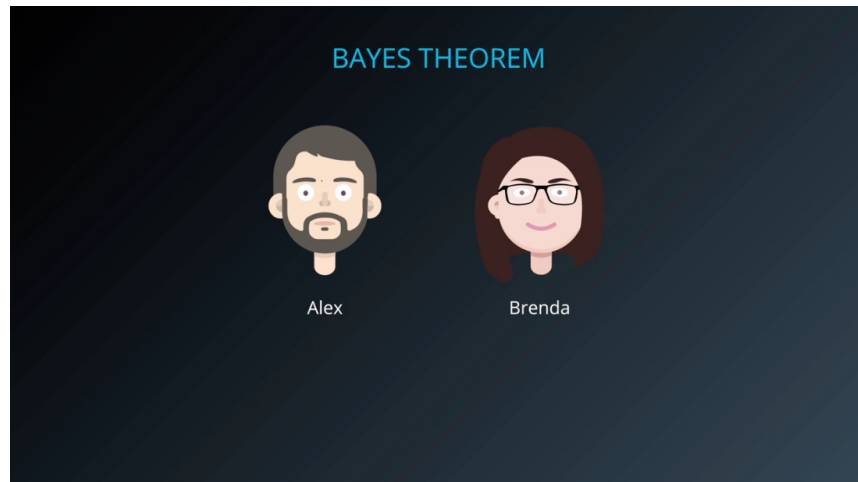
2024/2025

V. Bayésien Naïf

Classifieur Bayésien Naïf: Introduction

Mise en situation:

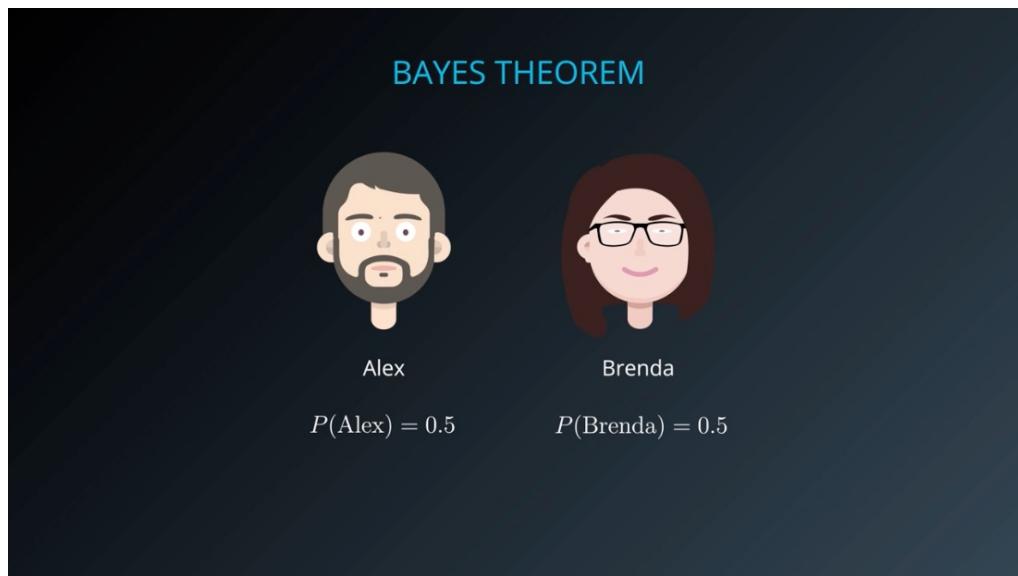
Alex et Brenda sont collègues et travaillent les mêmes jours. Un matin, vous apercevez brièvement quelqu'un passer au bureau. Vous ne pouvez pas dire avec certitude si c'était Alex ou Brenda. Quelle est la probabilité que la personne aperçue soit Alex ? Et quelle est la probabilité que ce soit Brenda ?



Classifieur Bayésien Naïf: Introduction

Mise en situation:

Alex et Brenda sont collègues et travaillent les mêmes jours. Un matin, vous apercevez brièvement quelqu'un passer au bureau. Vous ne pouvez pas dire avec certitude si c'était Alex ou Brenda. Quelle est la probabilité que la personne aperçue soit Alex ? Et quelle est la probabilité que ce soit Brenda ?



Classifieur Bayésien Naïf: Introduction

Mise en situation:

Alex et Brenda sont collègues et travaillent les mêmes jours. Un matin, vous apercevez brièvement quelqu'un passer au bureau. Vous ne pouvez pas dire avec certitude si c'était Alex ou Brenda. Quelle est la probabilité que la personne aperçue soit Alex ? Et quelle est la probabilité que ce soit Brenda ?

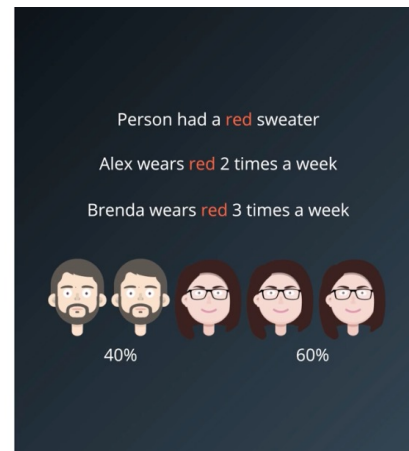
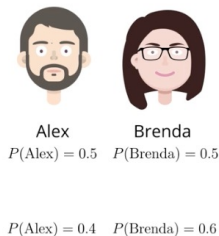
Infos supplémentaires:

La personne vue au bureau, porte un pull rouge. On sait que :

- Alex porte un pull rouge 2 fois par semaine
- Brenda porte un pull rouge 3 fois par semaine
- Les jours spécifiques où chacun porte un pull rouge ne sont pas connus

Quelle est la probabilité que la personne observée portant un pull rouge soit Brenda ?

BAYES THEOREM



Classifieur Bayésien Naïf: Introduction

BAYES THEOREM

connu

La probabilité que Alex
porte un pull rouge

La probabilité que Brenda
porte un pull rouge

déduit

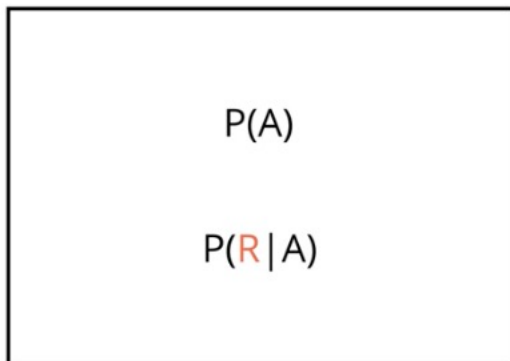
La probabilité que la
personne qui porte rouge
soit Alex

La probabilité que la
personne qui porte rouge
soit Brenda

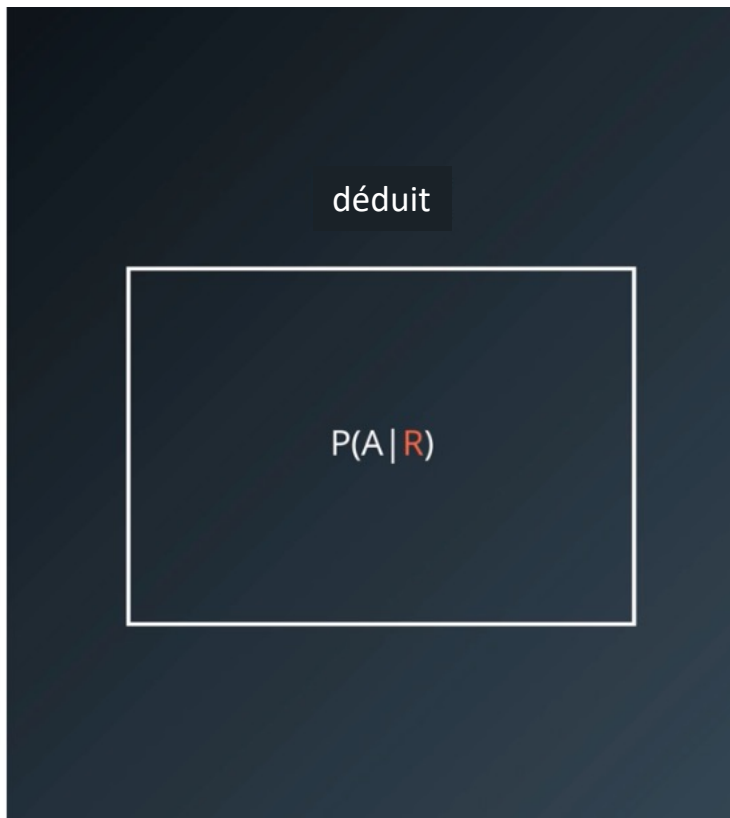
Classifieur Bayésien Naïf: Introduction

BAYES THEOREM

connu



déduit



Classifieur Bayésien Naïf: Introduction

Mise en situation:

Alex et Brenda sont collègues et travaillent les mêmes jours. Un matin, vous apercevez brièvement quelqu'un passer au bureau. Vous ne pouvez pas dire avec certitude si c'était Alex ou Brenda. Quelle est la probabilité que la personne aperçue soit Alex ? Et quelle est la probabilité que ce soit Brenda ?

Infos supplémentaires:

La personne vue au bureau, porte un pull rouge. On sait que :

- Alex porte un pull rouge 2 fois par semaine
- Brenda porte un pull rouge 3 fois par semaine
- Les jours spécifiques où chacun porte un pull rouge ne sont pas connus

Infos supplémentaires:

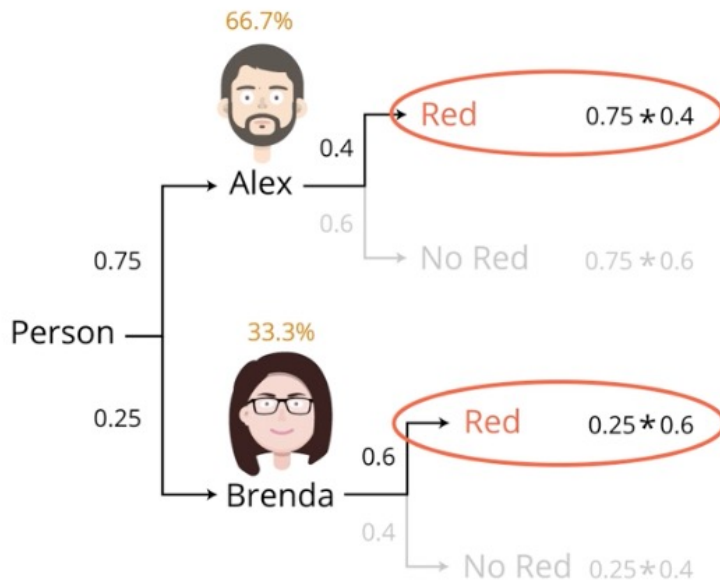
On sait que :

- Alex se rend au bureau 3 fois par semaine.
- Brenda se rend au bureau 1 fois par semaine.
- Les jours précis où chacun vient au bureau ne sont pas spécifiés.

Quelle est la probabilité que la personne observée portant un pull rouge soit Brenda ?

Classifieur Bayésien Naïf: Introduction

BAYES THEOREM



Alex se rend au bureau 3 fois par semaine
Brenda se rend au bureau 1 fois par semaine

La personne porte un pull rouge

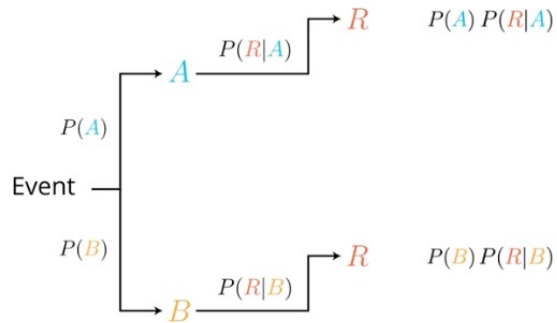
Alex porte un pull rouge 2 fois par semaine
Brenda porte un pull rouge 3 fois par semaine

$$P(A|R) = \frac{0.75 * 0.4}{0.75 * 0.4 + 0.25 * 0.6} = 0.67$$

$$P(B|R) = \frac{0.25 * 0.6}{0.75 * 0.4 + 0.25 * 0.6} = 0.33$$

Classifieur Bayésien Naïf: Théorème

BAYES THEOREM



$$P(A|R) = \frac{P(A)P(R|A)}{P(A)P(R|A) + P(B)P(R|B)}$$

$$P(B|R) = \frac{P(B)P(R|B)}{P(A)P(R|A) + P(B)P(R|B)}$$

Classifieur Bayésien Naïf: Quiz

3 emails **Spam**

Win **money** now!

Make cash **easy**!

Cheap **money**, reply

5 emails **Ham**

How are you?

There you are

Can I borrow **money**?

Say hi to grandma.

Was the exam **easy**?

Quelle est la probabilité qu'un e-mail contienne le mot 'easy', sachant qu'il s'agit d'un spam ?

- 1/5
- 1/4
- 1/3
- 2/3
- 4/5

Quelle est la probabilité qu'un e-mail contienne le mot 'money', sachant qu'il s'agit d'un spam ?

- 1/5
- 1/4
- 1/3
- 2/3
- 4/5

Classifieur Bayésien Naïf: Quiz

3 emails **Spam**

Win **money** now!

Make cash **easy**!

Cheap **money**, reply

5 emails **Ham**

How are you?

There you are

Can I borrow **money**?

Say hi to grandma.

Was the exam **easy**?

Quelle est la probabilité qu'un e-mail contienne le mot 'easy', sachant qu'il s'agit d'un spam ?

- 1/5
- 1/4
- **1/3**
- 2/3
- 4/5

Quelle est la probabilité qu'un e-mail contienne le mot 'money', sachant qu'il s'agit d'un spam ?

- 1/5
- 1/4
- 1/3
- **2/3**
- 4/5

Classifieur Bayésien Naïf: Quiz - correction

NAIVE BAYES



Spam

Win money now!
Make cash easy!
Cheap money, reply.

Ham

How are you?
There you are!
Can I borrow money?
Say hi to grandma.
Was the exam easy?

$P(\text{spam} \mid \text{'easy'})$

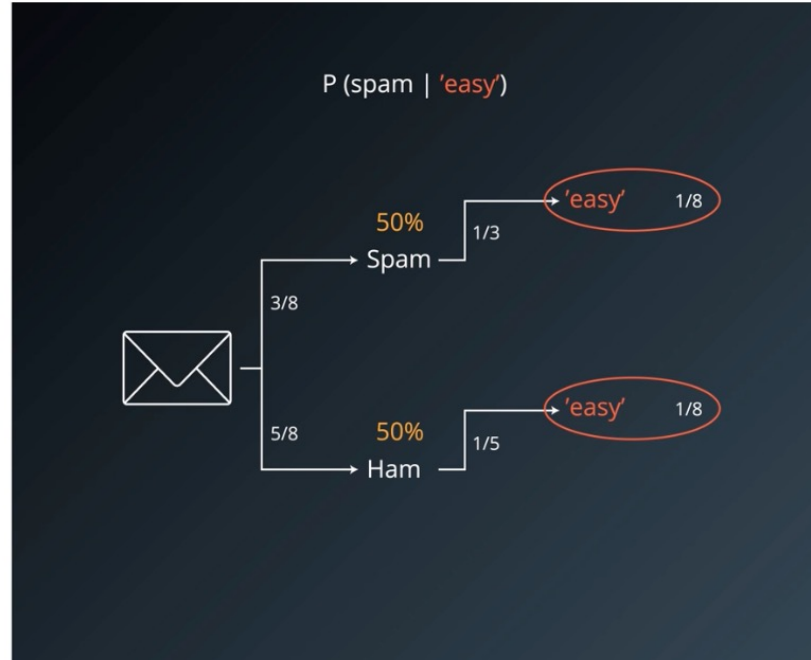


Classifieur Bayésien Naïf: Quiz - correction

NAIVE BAYES



Spam Win money now! Make cash easy! Cheap money, reply.
Ham How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?



Classifieur Bayésien Naïf: Quiz - correction

NAIVE BAYES



Spam Win money now! Make cash easy! Cheap money, reply.
Ham How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?

P (spam | 'money')

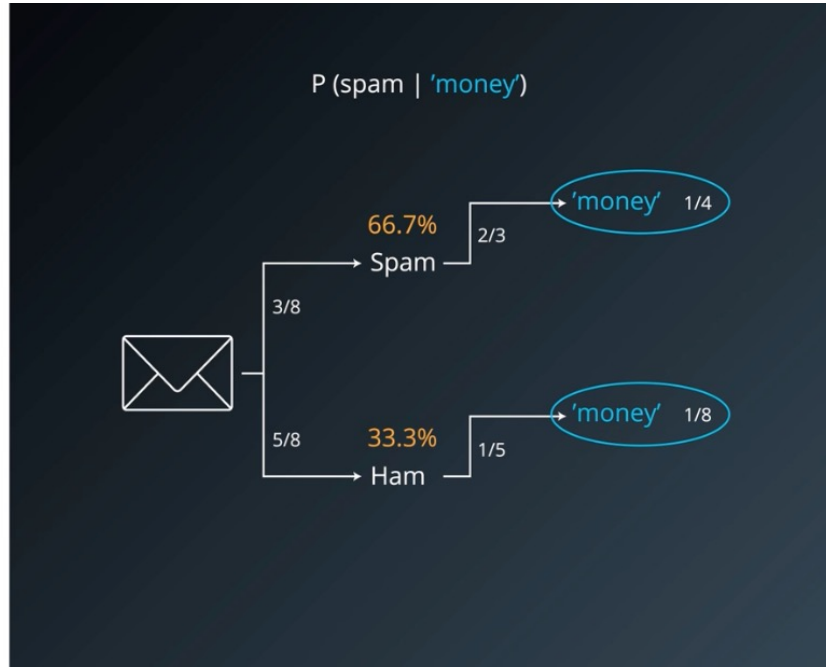


Classifieur Bayésien Naïf: Quiz - correction

NAIVE BAYES



Spam Win money now! Make cash easy! Cheap money, reply.
Ham How are you? There you are! Can I borrow money? Say hi to grandma. Was the exam easy?



Classifieur Bayésien Naïf: résumé

Définition: Algorithme de classification probabiliste basé sur le théorème de Bayes

Principe:

- Calcule la probabilité d'appartenance à chaque classe
- Sélectionne la classe avec la plus haute probabilité

Caractéristiques clés :

- Suppose l'indépendance entre les features (hypothèse "naïve")
- Rapide et efficace, surtout avec de grandes données
- Performant en classification de texte et filtrage de spam

Formule : $P(A|B) \propto P(B|A) * P(A)$

$$P(A \cap B) = P(A)P(B) \quad (\text{hypothèse "naïve"})$$

Avantages:

- Simple et rapide à implémenter
- Fonctionne bien avec peu de données d'entraînement
- Gère facilement les problèmes multi-classes

Limites:

- L'hypothèse d'indépendance est souvent irréaliste
- Performances limitées sur des données très corrélées

Applications courantes:

- Filtrage de spam
- Analyse de sentiments
- Catégorisation de documents

Classifieur Bayésien Naïf: Exercice

Projet pratique : Construire un classificateur de spam

La détection de spam est l'une des principales applications de l'apprentissage automatique sur Internet aujourd'hui. Pratiquement tous les grands fournisseurs de services de messagerie électronique disposent de systèmes de détection de spam intégrés et classent automatiquement ces courriers comme "Courrier indésirable".

Dans cet exercice, nous utiliserons l'algorithme Naive Bayes pour créer un modèle capable de classer les messages SMS d'un ensemble de données comme spam ou non spam, en fonction de l'entraînement que nous donnerons au modèle. Il est important d'avoir une certaine intuition de ce à quoi pourrait ressembler un message texte indésirable.

<https://github.com/elhidali/EPISEN-2024>