Classification Probabiliste
Naïve Bayes

Plan

Naive Bayes TL;DR I

- Naive Bayes est un algorithme de classification probabiliste.
- Étant donné une instance $\mathbf{x} = [x_1, \dots x_d]$ et un ensemble de valeurs cibles $y_1, \dots, y_k$,
- Il calculera toutes les probabilités conditionnelles :

$$P(y_i|\mathbf{x}), y_i \in \{y_1, \dots y_k\}$$

- et prédit la valeur cible y_i qui a la plus grande probabilité $P(y_i|\mathbf{x})$.

Naive Bayes TL;DR II

- Pour calculer $P(y_i|\mathbf{x})$ il utilise le théorème de Bayes

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)p(y_i)}{P(\mathbf{x})}$$

  et une hypothèse simplificatrice qui permet de calculer facilement $P(\mathbf{x}|y_i)$ comme suit :

$$P(\mathbf{x}|y_i) = P(x_1|y_i)P(x_2|y_i)\ldots P(x_d|y_i)$$

Naive Bayes, exemple TL;DR I

- Prenons l'exemple de Titanic
- La variable cible est le résultat et prend des valeurs dans $\{\text{mort}, \text{surv}\}$
- Une instance $\mathbf{x}$ a trois attributs prédictifs $\mathbf{x} = [\text{class}, \text{sex}, \text{age}]$

Naive Bayes, exemple TL;DR II

- Pour déterminer la valeur cible d'une instance donnée, par exemple $\mathbf{x} = [\text{first, female, enf}]$, naive Bayes calculera :

$$P(\text{mort}|\text{first, female, enf}) = \frac{P(\text{first, female, enf}|\text{mort})P(\text{mort})}{P(\text{first, female, enf})}$$
$$= \frac{P(\text{first}|\text{mort})P(\text{female}|\text{mort})P(\text{enf}|\text{mort})P(\text{mort})}{P(\text{first, female, enf})}$$

and

$$P(\text{surv}|\text{first, female, enf}) = \frac{P(\text{first, female, enf}|\text{surv})P(\text{surv})}{P(\text{first, female, enf})}$$
$$= \frac{P(\text{first}|\text{surv})P(\text{female}|\text{surv})P(\text{enf}|\text{surv})P(\text{surv})}{P(\text{first, female, enf})}$$

et sélectionnera la valeur cible ayant la plus grande probabilité.

- Nous savons déjà, grâce au TP1, comment calculer tous les $P(\text{first}|\text{surv}), P(\text{female}|\text{surv}), P(\text{enf}|\text{surv}), P(\text{surv})$

Plan

1. Naive Bayes TL;DR
2. **Théorème de Bayes et règle de Bayes**, passer à l'exemple simple de Bayes, slide 16
3. Naive Bayes

Introduction

- Avant de parler de naive Bayes, nous allons réviser le théorème de Bayes
- et discuter de la règle de Bayes
- qui constituent la base de naive Bayes
- Pour tout ce qui suit, nous n'avons besoin que des concepts (distributions de probabilités) du TP1.

Théorème de Bayes I

- Laissons Y être la variable cible qui prend les valeurs $y \in \mathcal{Y} = \{y_1, \ldots, y_k\}$
- Nous avons également une variable prédictive X qui prend les valeurs $x \in \mathcal{X}$
- Étant donné une observation x de la variable prédictive, nous aimerions calculer la probabilité qu'elle appartienne à la classe y :
  - $P(y|x)$ est la *probabilité postérieure* que x appartient à la classe y
  - $P(y)$ est la *probabilité a priori* de la classe y, c'est-à-dire avant l'observation de x.
- Rappel sur les probabilités conjointes et conditionnelles :

$$\begin{aligned} P(X,Y) &= P(Y|X)P(X) \\ &= P(X|Y)P(Y) \end{aligned}$$

Nous pouvons résoudre $P(Y|X)$, ce qui revient à théorème de Bayes

Théorème de Bayes et règle de Bayes I

- *Théorème de Bayes* : transforme la probabilité a priori en probabilité a posteriori à l'aide de l'information contenue dans l'instance x.

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$
$$= \frac{P(\mathbf{x}|y)P(y)}{\sum_{y_i \in \mathcal{Y}} P(\mathbf{x}|y_i)P(y_i)}$$

le dénominateur $P(\mathbf{x}) = \sum_{y_i \in \mathcal{Y}} P(\mathbf{x}|y_i)P(y_i)$ est là pour s'assurer que: $\sum_{y_i \in \mathcal{Y}} P(y'|\mathbf{x}) = 1$

Théorème de Bayes et règle de Bayes II

- *Règle de Bayes* : prédire comme classe y celle qui a la plus grande probabilité a posteriori, cad:

$$y_{map} = \underset{y \in \mathcal{Y}}{\operatorname{argmax}}\ P(y|\mathbf{x})$$

Exemple pratique de théorème de Bayes

Examinons le théorème de Baye à l'aide de l'ensemble de données du Titanic :

- X est le vecteur des variables prédictives, $X = [\text{class, age, sex}]$
- Y est la variable result qui prend des valeurs y dans $\mathcal{Y} = \{\text{mort, surv}\}$.
- Nous voulons, à partir d'une instance spécifique, par exemple

$$\mathbf{x} = [\text{crew, adu, enf}]$$

, prédire la valeur du result, cad, mort ou surv.

Exemple pratique de théorème de Bayes I

- Le théorème de Baye nous aidera à calculer la probabilité postérieure de la variable result compte tenu d'une instance, cad:

$$P(\text{result}|\text{class, sex, age}) = \frac{P(\text{class, sex, age}|\text{result})P(\text{result})}{P(\text{class, sex, age})}$$

Exemple pratique de théorème de Bayes II

- La variable result prend des valeurs dans $\{$mort, surv$\}$.
- Nous devons donc calculer les deux probabilités a posteriori :

$$P(\text{result=mort}|\text{class, sex, age}) = \frac{P(\text{class, sex, age}|\text{result=mort})P(\text{result=mort})}{P(\text{class, sex, age})}$$

$$P(\text{result=surv}|\text{class, sex, age}) = \frac{P(\text{class, sex, age}|\text{result=surv})P(\text{result=surv})}{P(\text{class, sex, age})}$$

Exemple pratique de règle de Bayes

- Après avoir calculé les deux probabilités, nous prédirons comme valeur de result pour notre instance celle qui a la plus grande probabilité, cad:

$$\text{result} := \underset{\text{mort, surv}}{\operatorname{argmax}}\{P(\text{result=mort}|\text{class, sex, age}), P(\text{result=surv}|\text{class, sex, age})\}$$

Exemple pratique simplifié du théorème de Bayes I

- Pour nous simplifier la vie, nous ne considérerons pour l'instant qu'une seule variable prédictive, par exemple sex :
- et le théorème de Bayes nous dit:

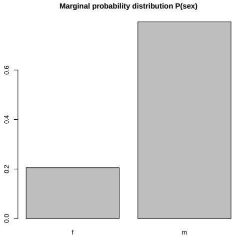
$$p(\text{result}|\text{sex}) = \frac{p(\text{sex}|\text{result})p(\text{result})}{p(\text{sex})} \tag{1}$$

- Dans le premier TP, nous avons calculé un certain nombre de distributions qui apparaissent dans eq 1
- Nous verrons plus tard comment naive Bayes traite plus d'une variable à la fois dans $\mathbf{x}$,

Distributions marginales, $P(\text{result})$, $P(\text{sex})$

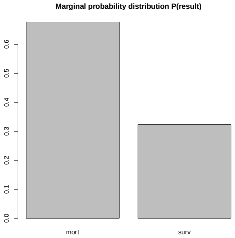
La marginale $P(\text{sex})$

```
        f         m
0.2053612 0.7946388
```



La marginal $P(\text{result})$

```
    mort     surv
0.676965 0.323035
```

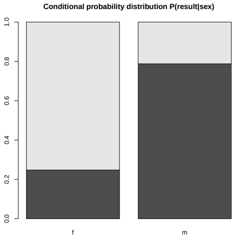


Il s'agit de la distribution a priori de la cible, qui apparaît dans la partie droite du théorème de Bayes eq 1.

Distribution conditionnelle, $P(\text{result}|\text{sex})$ I

Distribution conditionnelle: $P(\text{result}|\text{sex})$

```
   result
sex        mort      surv
  f 0.2477876 0.7522124
  m 0.7878788 0.2121212
```



- La somme des probabilités est de 1.0 par ligne (nous fixons la variable du sexe).

Il s'agit de la distribution postérieure qui apparaît dans la partie gauche du théorème de Bayes, eq 1

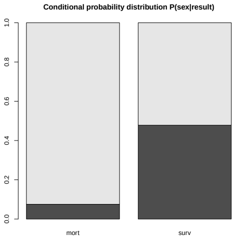
Distribution conditionnelle, $P(\text{result}|\text{sex})$ II

- Nous avons utilisé la distribution conditionnelle de la variable cible (result) en fonction de chaque attribut pour sélectionner les attributs les plus informatifs.
- En fait, si nous n'avons qu'un seul attribut prédictif, le sex, la distribution conditionnelle $P(\text{résultat}|\text{sexe})$ est tout ce dont nous avons besoin pour sélectionner la classe la plus probable pour une instance.
- Mais nous en avons plus (donc naive Bayes)

Distribution conditionnelle, $P(\text{sex}|\text{result})$

Il existe une autre distribution conditionnelle: $P(\text{sex}|\text{result})$

```
    result
sex        mort       surv
  f 0.07516779 0.47819972
  m 0.92483221 0.52180028
```



- La somme des probabilités est de 1.0 par colonne (nous fixons la variable result).

nous n'avons pas calculé cette distribution dans le premier TP, mais elle apparaît dans le théorème de Bayes et nous l'utiliserons dans le cadre de Bayes naïf.

Théorème de Bayes sur le Titanic avec les variables result et sex. I

- Le théorème de Bayes nous indique que nous pouvons calculer le $P(\text{result}|\text{sex})$ comme suit

$$P(\text{result}|\text{sex}) = \frac{P(\text{sex}|\text{result})P(\text{result})}{P(\text{sex})}$$

- Calculons donc la probabilité de mort pour une femme, cad P(result=mort|sex=f):

$$P(\text{result=mort}|\text{sex=f}) = \frac{P(\text{sex=f}|\text{result=mort})P(\text{result=mort})}{P(\text{sex=f})}$$

Théorème de Bayes sur le Titanic avec les variables result et sex. II

- en consultant les distributions

  P(sex|result) (slide 20), P(result), P(sex) (slide 17)

  en choisissant les probabilités appropriées et en remplaçant ce qui précède, on obtient:

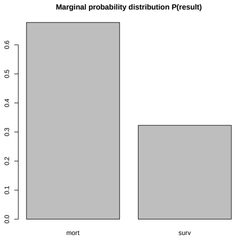
$$P(\text{result=mort}|\text{sex=f}) = \frac{0.07516779 * 0.676965}{0.2053612} = 0.2477876$$

ce qui est bien la valeur correcte de $P(\text{result=mort}|\text{sex=f})$ comme nous pouvons le voir dans $P(\text{result}|\text{sex})$ (slide 18)

Passez directement à Naive Bayes, slide 27, ou continuez avec l'intuition.

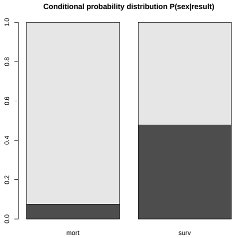
Théorème de Bayes sur titanic, intuition I

- Il y a une façon intéressante de penser à ce que fait Bayes.
- Avant d'observer l'instance $\mathbf{x}$, la seule chose que nous puissions dire à propos de la variable result provient de sa distribution a priori $P(\text{result})$.



Théorème de Bayes sur titanic, intuition II

- dès que nous observons notre instance, sex $= f$, nous utilisons $P(\text{sex} = f|\text{result})$:

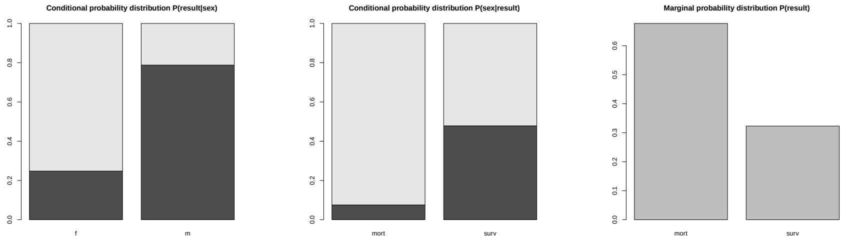


pour ajuster les probabilités de result données par l'a priori $P(\text{result})$

Théorème de Bayes sur titanic, intuition III

- en particulier:

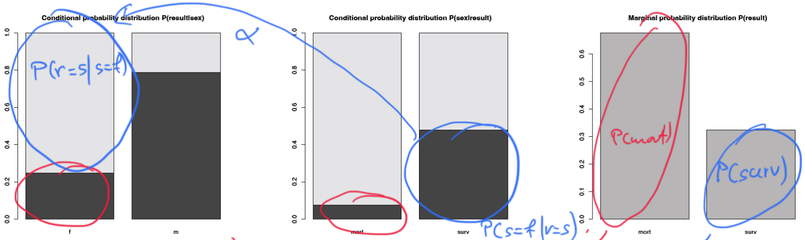
$$P(\text{result}|\text{sex=f}) \quad \propto \quad P(\text{sex} = f|\text{result}) \quad \times \quad P(\text{result})$$



Théorème de Bayes sur titanic, intuition IV

- en particulier:

$$P(\text{result}|\text{sex=f}) \quad \propto \quad P(\text{sex} = f|\text{result}) \quad \times \quad P(\text{result})$$



$P(\text{result}=m|\text{sex}=f) \propto P(\text{sex}=f|\text{result}=m) \times$

Plan

De Bayes à naive Bayes

- Jusqu'à présent, nos instances $\mathbf{x}$ n'avaient qu'une seule variable, sex.
- Comment pouvons-nous faire la même prédiction lorsque nous avons plus d'une variable, par exemple comme dans Titanic où $\mathbf{x} = [\text{class, sex, age}]$, cad comment pouvons-nous calculer :

$$P(\text{result}|\text{class, sex, age}) = \frac{P(\text{class, sex, age}|\text{result})P(\text{result})}{P(\text{class, sex, age})}$$

c'est ici que naive Bayes entre en jeu.

Naive Bayes

- Voyons donc comment nous allons calculer

$$P(\text{result}|\text{class, sex, age}) = \frac{P(\text{class, sex, age}|\text{result})P(\text{result})}{P(\text{class, sex, age})}$$

  avec l'aide de naive Bayes.

- Le principal problème est de savoir comment obtenir:

$$P(\text{class, sex, age}|\text{result})$$

- nous savons comment obtenir

$$P(\text{class}|\text{result}), \quad P(\text{sex}|\text{result}), \quad P(\text{age}|\text{result})$$

  mais pas comment obtenir $P(\text{class, sex, age}|\text{result})$

Naive Bayes, obtenant $P(\text{class, sex, age}|\text{result})$

- Pour calculer $P(\text{class, sex, age}|\text{result})$ Naive Bayes fait une hypothèse simplificatrice
- Il suppose que les attributs prédictifs sont indépendants de l'attribut cible
- Cette hypothèse nous permet de calculer :

$$P(\text{class, sex, age}|\text{result}) = P(\text{class}|\text{result})P(\text{sex}|\text{result})P(\text{age}|\text{result})$$

et nous savons comment calculer toutes les distributions de ce produit.

Naive Bayes, obtenant $P(\text{class, sex, age})$

- Le dernier élément qui nous manque est le dénominateur, cad $P(\text{class, sex, age})$.
- Nous pouvons facilement l'obtenir sous la forme suivante:

$$\begin{aligned} P(\text{class, sex, age}) = & \; P(\text{class, sex, age}|\text{result=mort})P(\text{result=mort}) \\ & + P(\text{class, sex, age}|\text{result=surv})P(\text{result=surv}) \end{aligned}$$

- Notez que le calcul de $P(\text{class, sex, age})$ n'est pas nécessaire si nous voulons trouver la valeur cible la plus probable, il n'est nécessaire que si l'on veut calculer la probabilité postérieure exacte.

Naive Bayes Model

- Étant donné une instance $\mathbf{x} = [x_1, \ldots x_d]$,
- et un ensemble de valeurs cibles$\mathcal{Y} = \{y_1, \ldots, y_k\}$
- Bayes classifiera $\mathbf{x}$ dans la classe la plus probable, comme suit:

$$y_{\mathsf{MAP}} = \underset{y_i \in \mathcal{Y}}{\operatorname{argmax}} \frac{P(\mathbf{x}|y_i)P(y_i)}{P(\mathbf{x})}$$

$$= \underset{y_i \in \mathcal{Y}}{\operatorname{argmax}} \frac{P(x_1, \ldots x_d|y_i)P(y_i)}{P(\mathbf{x})}$$

qui, sous l'hypothèse simplificatrice de Naive Bayes, est donnée par

$$= \underset{y_i \in \mathcal{Y}}{\operatorname{argmax}} \frac{P(x_1|y_i)P(x_2|y_i)\ldots P(x_d|y_i)P(y_i)}{P(\mathbf{x})}$$

Naive Bayes, entraînement et test

Entraînement:

- Estimer $P(Y) = \{P(y_1), \ldots, P(y_k\}$, cad la probabilité de chaque valeur cible.
- Pour chaque attribut prédictif X_i
  - Estimer la distribution conditionnelle : $P(X_i|Y)$.

Test:

- Pour une instance donnée $\mathbf{x}$
  - Estimer:

$$P(Y|\mathbf{x}) = \{P(y_1|\mathbf{x}), \ldots P(y_k|\mathbf{x}\}$$

  - Prédire la valeur cible avec la plus grande probabilité:

$$\text{prediction} = \underset{y_i}{\operatorname{argmax}}\{P(y_1|\mathbf{x}), \ldots P(y_k|\mathbf{x}\}$$

Estimation de la distribution a priori de variable cible: $P(Y)$ I

- Peut être facilement réalisé à partir de l'ensemble d'apprentissage (par exemple, à l'aide de prop.table)
- Par exemple, dans le cas de Titanic il nous faut:

$$P(\text{result}) = \{P(\text{result} = \text{mort}), P(\text{result} = \text{surv})\} \tag{2}$$

```
Computed by Naive Bayes:

A-priori probabilities:
Y
    mort     surv
0.676965 0.323035     =>  P(mort)  P(surv)
```

Estimation de la distribution conditionnelle: $P(X_i|Y)$, X_i qualitatif I

- Peut être facilement réalisé à partir de l'ensemble d'apprentissage (par exemple, à l'aide de prop.table)
- Par exemple, dans le cas de Titanic, nous avons besoin les:

$$P(\text{class}|\text{result}), P(\text{age}|\text{result}), P(\text{sex}|\text{result}) \tag{3}$$

```
Computed by Naive Bayes:

Conditional probabilities:
      class                                                 => P(class | result)
Y            crew      first     second      third
  mort 0.44383562 0.08082192 0.10958904 0.36575342  => P(crew|mort) P(first|mort) ...
  surv 0.30000000 0.29459459 0.14864865 0.25675676  => P(crew|surv) P(first|surv) ...

      age
Y             adu        enf
  mort 0.96027397 0.03972603   => P(adu|mort) P(enf|mort)
  surv 0.91351351 0.08648649   => P(adu|surv) P(enf|surv)

      sex
Y               f          m
  mort 0.07671233 0.92328767   => P(f|mort) P(m|mort)
  surv 0.47567568 0.52432432   => P(f|surv) P(m|surv)
```

Estimation de $P(Y|\mathbf{x})$ I

- Au moment du test, étant donné une instance $\mathbf{x}$, nous utiliserons directement:
  - les tableaux de probabilite conditionneles $P(X_i|Y)$ pour toutes les attributes X_i
  - ainsi que $P(Y)$

  pour calculer la probabilite de chaque valeur de variable cible cad la distribution $P(Y|\mathbf{x})$.

Estimation de $P(Y|\mathbf{x})$ II

- Par exemple, pour l'instance de titanic suivante:

```
class age sex
first adu   m
```

- nous allons calculer:

$$P(\text{result=mort}|\text{first, m, adu }) = \frac{P(\text{first, m, adu }|\text{result=mort})P(\text{result=mort})}{P(\text{first, m, adu })}$$

$$P(\text{result=surv}|\text{first, m, adu }) = \frac{P(\text{first, m, adu }|\text{result=surv})P(\text{result=surv})}{P(\text{first, m, adu })}$$

- On peut directement recupperer les deux probabilites

$$P(\text{result=mort}) \text{ et } P(\text{result=surv})$$

depuis le tableau qui ce trouve au dessous d' equation 2

Estimation de $P(Y|\mathbf{x})$ III

- On peut calculer

$$P(\text{first, m, adu }|\text{result=mort}) \text{ et } P(\text{first, m, adu }|\text{result=surv})$$

comme suit (hypothese simplificatrice):

$$P(\text{first, m, adu }|\text{res=mort}) = P(\text{first}|\text{res=mort})P(\text{m}|\text{res=mort})P(\text{adu}|\text{res=mort})$$
$$P(\text{first, m, adu }|\text{res=surv}) = P(\text{first}|\text{res=surv})P(\text{m}|\text{res=surv})P(\text{adu}|\text{res=surv})$$

On peut directement reccuperer les probabilites:

$$P(\text{first}|\text{res=mort}), P(\text{m}|\text{res=mort}), P(\text{adu}|\text{res=mort})$$
$$P(\text{first}|\text{res=surv}), P(\text{m}|\text{res=surv}), P(\text{adu}|\text{res=surv})$$

depuis le table qui se trouve au dessous de equation 3

Estimation de $P(Y|\mathbf{x})$ IV

- Dernier chose a calculer: $P(\text{first, m, adu })$, facile à calculer

$$\begin{aligned} P(\text{first, m, adu }) = & P(\text{first, m, adu }|\text{result=mort})P(\text{result=mort}) + \\ & P(\text{first, m, adu }|\text{result=surv})P(\text{result=surv}) \end{aligned}$$

  on vient de voir comment calculer
  $P(\text{first, m, adu }|\text{result=mort})P(\text{result=mort})$ et
  $P(\text{first, m, adu }|\text{result=surv})P(\text{result=surv})$
- On a maintenant les éléments qui il nous faut pour calculer:

$$P(\text{result=mort}|\text{first, m, adu }) \text{ et } P(\text{result=surv}|\text{first, m, adu })$$

  et en suite choisir la valeur la plus probable.

Estimation de la distribution conditionnelle : $P(X_i|Y)$, X_i quantitatif I

- Pour les attributs continus, nous supposerons qu'ils suivent une distribution normale compte tenu de la variable cible Y, cad:

$$P(X_i|Y) = \mathcal{N}(X_i; \mu_Y, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\frac{(X_i-\mu_y)^2}{\sigma_Y^2}}$$

- pour chaque attribut continu, nous calculerons sa moyenne conditionnelle et son écart-type pour chaque valeur de la variable cible Y.

Estimation de la distribution conditionnelle : $P(X_i|Y)$, X_i quantitatif II

- Par exemple, pour l'iris, nous aurons douze distributions normales (quatre attributs $\times$ trois classes).

$$N(X_i; \mu_{X_i,\text{setosa}}, \sigma^2_{X_i,\text{setosa}}),$$
$$N(X_i; \mu_{X_i,\text{virginica}}, \sigma^2_{X_i,\text{virginica}}),$$
$$N(X_i; \mu_{X_i,\text{virginica}}, \sigma^2_{X_i,\text{virginica}})$$

où X_i correspond à l'un des quatre attributs prédictifs de l'iris, c'est-à-dire :

$$X_i \in \{\text{sepal.length, sepal.width, petal.length, petal.width}\}$$

- Pendant l'apprentissage, nous calculons la moyenne conditionnelle et l'écart-type pour chaque classe.

Estimation de la distribution conditionnelle : $P(X_i|Y)$, X_i quantitatif III

- Pour l'iris, cela se traduira par les tableaux suivants :

```
Conditional probabilities:
                 sepal_length
Y                     [,1]      [,2]  => column[,1] is the conditional mean of sepal_length for each c
  Iris_setosa     5.065385 0.3224187     column[,2] is the standard deviation of sepal_length for each
  Iris_versicolor 5.844444 0.4676729
  Iris_virginica  6.663636 0.6366171

                 sepal_width
Y                     [,1]      [,2] => column[,1] is the conditional mean of sepal_width for each cla
  Iris_setosa     3.542308 0.3360632     column[,2] is the standard deviation of sepal_width for each
  Iris_versicolor 2.722222 0.2606697
  Iris_virginica  2.959091 0.3620827

                 petal_length
Y                     [,1]      [,2] => ...
  Iris_setosa     1.469231 0.1975231
  Iris_versicolor 4.203704 0.4879167
  Iris_virginica  5.586364 0.6057624

                 petal_width
Y                      [,1]      [,2] => ...
  Iris_setosa     0.2807692 0.1166850
  Iris_versicolor 1.3037037 0.1764642
  Iris_virginica  1.9863636 0.3028365
```

Estimation de la distribution conditionnelle : $P(X_i|Y)$, X_i quantitatif IV

- Au moment du test, étant donné une instance $\mathbf{x}$, nous utiliserons les moyennes conditionnelles et les écarts types pour calculer toutes les probabilités $P(X_i|Y)$ avec la distribution normale.
- Par exemple, pour l'instance d'iris suivante :

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
5.1	3.5	1.4	0.2

nous devons calculer:

$$
\begin{aligned}
P(\text{Sepal.length} = 5.1|\text{type} = \text{setosa}) &= \mathcal{N}(5.1; 5.065385, 0.3224187)\\
P(\text{Sepal.length} = 5.1|\text{type} = \text{versicolor}) &= \mathcal{N}(5.1; 5.844444, 0.4676729)\\
P(\text{Sepal.length} = 5.1|\text{type} = \text{virginica}) &= \mathcal{N}(5.1; 6.663636, 0.6366171)\\
&\dots\\
P(\text{Petal.wifth} = 0.2|\text{type} = \text{virginica}) &= \mathcal{N}(0.2; 1.9863636, 0.3028365)
\end{aligned}
$$

Estimation de la distribution conditionnelle : $P(X_i|Y)$, X_i quantitatif V

- que nous utiliserons pour calculer les probabilités postérieures:

  $P(\text{type} = \text{setosa}|\mathbf{x}), P(\text{type} = \text{virginica}|\mathbf{x}), P(\text{type} = \text{versicolor}|\mathbf{x})$

- uis choisir le type qui a la plus grande probabilité.

Bilan

- Avantages :
  - méthode très simple aux performances honnêtes. Taux de réussite remarquable dans certains domaines (ex. recherche d'information)
  - intrinsèquement robuste aux valeurs manquantes
  - robuste aux variables non pertinentes
  - transparence de la méthode appréciée dans certains domaines (ex. médecine)
  - marche bien même lorsque l'indépendance conditionnelle des variables prédictives non vérifíee
- Inconvénients
  - construit des modèles simples qui peuvent être trop rigides pour des problèmes complexes
  - pour les variables continues, le problème d'estimation de densité reste entier