Université Constantine 2
جامعة قسنطينة 2

Apprentissage machine 1

Chapitre 4 : Classification supervisée

Classificateur naif bayes

Ouadfel Salima
Faculté NTIC/IFA
salima.ouadfel@univ-constantine2.dz

Apprentissage machine 1

Chapitre 3 : Classification supervisée

Classificateur naif bayes

Faculté NTIC/IFA

salima.ouadfel@univ-constantine2.dz

Etudiants concernés

Faculté/Institut	Département	Niveau	Spécialité
Nouvelles technologies	IFA	Master1	STIC

Classificateur Naif Bayes

Le classificateur Naif Bayes est une méthode de classification supervisée.

Le classificateur Naif Bayes se base sur le théorème de Bayes fondé sur les probabilités conditionnelles $P(Y/X)$.

Probabilités conditionnelles $P(Y/X)$: Quelle est la probabilité qu'un événement Y se produise sachant qu'un autre événement X s'est déjà produit?

Classificateur Naif Bayes

Soit un jeu donnée $X = \left((X_1, Y_1), (X_2, Y_2) \ldots (X_n, X_n)\right)$ constitué de n paires(X_i, Y_i), telle que $X_i = (X_{i1}, X_{i2} \ldots X_{ip})$ l'ensemble de p attributs (variables explicatives) et Y_i la variable à prédire (l'attribut classe ou label). Si le jeu de donnée X peut être classer en K classes alors $Y_i \in \{1,2, \ldots . K\}$

Le but est de prédire l'étiquette Y_i associée à une nouvelle entrée X_i

On cherche à estimer $p(Y_i = j / X_i = x)$ qui représente la probabilité que la donnée X_i soit affectée à la classe C_j $(Y_i = j)$ si $X_i = x$.

Classificateur Naif Bayes

	X_1	X_2	X_3	X_4	Y
Jour	Temps	Température	Humidité	Vent	Jouer
1	Ensoleillé	chaude	Élevée	faible	non
2	Ensoleillé	chaude	Élevée	fort	non
3	Couvert	chaude	Élevée	faible	oui
4	pluie	douce	Élevée	faible	oui
5	pluie	fraiche	normale	faible	oui
6	pluie	fraiche	normale	fort	non
7	Couvert	fraiche	normale	fort	oui
8	Ensoleillé	douce	Élevée	faible	non
9	Ensoleillé	fraiche	normale	faible	oui
10	Pluie	douce	normale	faible	Oui
11	Ensoleillé	douce	normale	Fort	Oui
12	Couvert	douce	Élevée	Fort	Oui
13	Couvert	chaude	normale	Faible	Oui
14	pluie	douce	Élevée	fort	Non

2 classes

$C_1 = Oui$

$C_2 = non$

$si\ (X_i = x = (Pluie, douce, élevée\ , Faible)$

quelle est la probabilité $p(Y_i = oui\ /X_i = x)$

Classificateur Naif Bayes

Selon la règle de bayes:

$$p(Y_i = j\ /X_i = x) = \frac{p(X_i = x/Y_i = j\ ) * p(Y_i = j)}{p(X_i = x)}$$

$p(Y_i = j\ /X_i = x)$: représente la probabilité conditionnelle de $Y_i = j$ sachant que $X_i = x$ C'est la probabilité a posteriori

$p(X_i = x/Y_i = j\ )$: représente la probabilité conditionnelle de $X_i = x$ si l'étiquette de X_i est$j$ $(Y_i = j)$ C'est la probabilité de vraisemblance

$p(X_i = x)$ représente la probabilité que $X_i = x$

C'est la probabilité a priori ou evidence

$p(Y_i = j)$ représente la probabilité que $Y_i = j$ **(c'est la proportion de données qui forment classe C_j)** C'est la probabilité a priori

Classificateur Naif Bayes

$$p(Y_i = j\ / X_i = x) = \frac{p(X_i = x / Y_i = j\ ) * p(Y_i = j)}{p(X_i = x)}$$

Le classifieur bayes naif se base sur **l'Hypothèse d'indépendance conditionnelle**

$$p(X_i = x / Y_i = j\ ) = p\big(X_i = (X_{i1}, X_{i2} \dots X_{ip}) / Y_i = j\ \big)$$

$$\boxed{\begin{aligned} p(X_i = x / Y_i = j\ ) = \ & p(X_i = X_{i1}\ / Y_i = j\ ) \times \\ & p(X_i = X_{i2}\ / Y_i = j\ ) \times \\ & \dots\dots \\ & p\big(X_i = X_{ip}\ / Y_i = j\ \big) \end{aligned}}$$

<u>**Estimée à l'aide des données d'apprentissage**</u>

Classificateur Naif Bayes

$$p(Y_i = j\,/X_i = x) = \frac{p(X_i = x/Y_i = j\,) * p(Y_i = j)}{p(X_i = x)}$$

La probabilité pour que la donnée X_i prenne la valeur x s'écrit:

$$p(X_i = x) = \sum_{k=1}^{K} p(X_i = x \; et \; Y_i = k)$$

$$\boxed{p(X_i = x) = \sum_{k=1}^{K} p(X_i = x/Y_i = k) \times p(Y_i = \mathbf{k})}$$

Classificateur Naif Bayes

$$p(Y_i = j / X_i = x) = \frac{p(X_i = x / Y_i = j) * p(Y_i = j)}{p(X_i = x)}$$

S’écrit:

$$p(Y_i = j / X_i = x) = \frac{\prod_{p=1}^{P} p(X_i = x_{ip} / Y_i = j) \times p(Y_i = j)}{\sum_{k=1}^{K} p(X_i = x / Y_i = k) \times p(Y_i = \mathrm{k})}$$

Classificateur Naif Bayes

On calcule la probabilité $p(Y_i = j\ /X_i = x)$ à chaque classe $C_j\ j = 1..K$ pour la donnée $X_i = x$, on décide de la classe d'appartenance de X_i en prenant le **maximum de probabilité a posteriori (maximum de vraisemblance)**:

$$Y_{MAP} = argmax\ _j\{p(Y_i = j/X_i = x)\}$$

$$Y_{MAP} = argmax\ _j\{\frac{p(X_i=x/Y_i=j\ )*p(Y_i=j)}{p(X_i=x)}\}$$

$$\boxed{Y_{MAP} = argmax_j\{p(X_i = x/Y_i= j\ ) * p(Y_i = j)\}}$$

Cas de variables prédictives qualitatives

Classificateur Naif Bayes

Exemple

On cherche le modèle de classification supervisée basé sur le classificateur Naif bayesian

Jour	Temps	Température	Humidité	Vent	Jouer
1	Ensoleillé	chaude	Élevée	faible	non
2	Ensoleillé	chaude	Élevée	fort	non
3	Couvert	chaude	Élevée	faible	oui
4	pluie	douce	Élevée	faible	oui
5	pluie	fraiche	normale	faible	oui
6	pluie	fraiche	normale	fort	non
7	Couvert	fraiche	normale	fort	oui
8	Ensoleillé	douce	Élevée	faible	non
9	Ensoleillé	fraiche	normale	faible	oui
10	Pluie	douce	normale	faible	Oui
11	Ensoleillé	douce	normale	Fort	Oui
12	Couvert	douce	Élevée	Fort	Oui
13	Couvert	chaude	normale	Faible	Oui
14	pluie	douce	Élevée	fort	Non

Exemple

X=Attribut	Valeurs de X	Y=Jouer = Oui	Y=Jouer = Non
Temps	Ensoleillé	2	3
	Couvert	4	0
	Pluie	3	2
Température	Chaude	2	2
	Douce	4	2
	Fraiche	3	1
Humidité	Elevée	3	4
	Normale	6	1
Vent	Fort	3	3
	Faible	6	2
Jouer		9	5

Classificateur Naif Bayes

Exemple

$$p(Y_i = j / X_i = x) = \frac{\prod_{p=1}^{P} p(X_i = x_{ip} / Y_i = j) \times p(Y_i = j)}{\sum_{k=1}^{K} p(X_i = x / Y_i = k) \times p(Y_i = k)}$$

X=Attribut	x=Valeurs	P(Y=Jouer = Oui)=9/14	P(Y=Jouer = Non)=5/14
Temps	Ensoleillé	P(Ensoleillé/jouer=oui)=2/9	P(Ensoleillé/jouer)non)=3/5
	Couvert	P(Couvert//jouer=oui)=4/9	P(Couvert//jouer=non)=0/5
	Pluie	P(Pluie/jouer=oui)=3/9	P(Pluie/jouer=non)=2/5
Température	Chaude	P(Chaude/jouer=oui)=2/9	P(Chaude/jouer=non)=2/5
	Tiède	P(Douce/jouer=oui)=4/9	P(Douce/jouer=non)=2/5
	Fraiche	P(Fraiche/jouer=oui)=3/9	P(Fraiche/jouer=non)=1/5
Humidité	Elevée	P(Elevée/jouer=oui)=3/9	P(Elevée/jouer=non)=4/5
	Normale	P(Normale/jouer=oui)=6/9	P(Normale/jouer=non)=1/5
Vent	Fort	P(Fort/jouer=oui)=3/9	P(Fort/jouer=non)=3/5
	Faible	P(Faible/jouer=oui)=6/9	P(Faible/jouer)non)=2/5
Jouer		9	5

Exemple

x = (ensoleillé, fraiche, élevée, fort), y Є {oui, non} .
On voudrait prédire la classe de la donnée X=x

$$Y_{MAP} = argmax \ \{p(Y = oui/X = x) , p(Y = non/X = x)\}$$

$$p(Y = oui\ /X = x) = p(X = x/Y = oui\ ) \times p(Y = oui)$$

$$p\big(Y = oui\ /X = (ensoleillé, fraiche, élevée, fort)\big) = p(X = (ensoleillé, fraiche, élevée, fort)/Y = oui\ ) \times p(Y = oui)$$

$$p\big(Y = oui\ /X = (ensoleillé, fraiche, élevée, fort)\big) = p(X = ensoleillé/Y = oui\ ) \times p(X = fraiche/Y = oui\ ) \times p(X = élevée/Y = oui\ ) \times p(X = fort/Y = oui\ )$$

$$p\big(Y = oui\ /X = (ensoleillé, fraiche, élevée, fort)\big) = \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} = \frac{2}{243}$$

$$p(Y = oui) = \frac{9}{14}$$

Exemple

x = (e$nsoleill$é, $fraiche$, élevée, $fort$), y ∈ {oui, non} .
On voudrait prédire la classe de la donnée X=x

$$Y_{MAP} = argmax \ \{p(Y = oui/X = x)\ , p(Y = non/X = x)\}$$

$p(Y = non\ /X = x)$= $p(X = x/Y = non) \times p(Y = non)$

$p\big(Y = non\ /X = (ensoleillé, fraiche, élevée, fort)\big)$ = $p(X = (ensoleillé, fraiche, élevée, fort)/Y = non\ ) \times$ $p(Y = non)$

$p\big(Y = non/X = (ensoleillé, fraiche, élevée, fort)\big)$ = $p(X = ensoleillé/Y = non) \times$
$p(X = fraiche/Y = non) \times$
$p(X = élevée/Y = non) \times$
$p(X = fort/Y = non)$

$p\big(Y = non/X = (ensoleillé, fraiche, élevée, fort)\big)$ = $\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} = \frac{36}{625}$

$p(Y = non)$= $\frac{5}{14}$

Exemple

x = (e*nsoleillé*, *fraiche*, *élevée*, *fort*), y Є {oui, non} .
On voudrait prédire la classe de la donnée X=x

$$Y_{MAP} = argmax \ \{p(Y = oui/X = x), p(Y = non/X = x)\}$$

$p(Y = oui/X = x)$= $p(X = x/Y = oui) \times p(Y = oui) = \frac{2}{243} * \frac{9}{14}$=$0{,}0053$
$p(Y = non\ /X = x)$= $p(X = x/Y = non) \times p(Y = non) = \frac{36}{625} * \frac{5}{14}$=0,020

Y_{MAP} =0,020

la donnée X=x sera classée dans la classe non

Estimateur de Laplace

Dans certains cas, la probabilité d'apparence d'une classe est nulle car il y a un manque d'exemples de cette classe. Afin de résoudre ce problème, on utilise **l'estimateur de Laplace qui consiste à ajouter une valeur** μ à chaque dénominateur pour un attribut donné,

Estimateur de Laplace

Dans certains cas, la probabilité d'apparence d'une classe est nulle car il y a un manque d'exemples de cette classe. Afin de résoudre ce problème, on utilise l'estimateur de Laplace qui consiste à ajouter une valeur μ à chaque dénominateur pour un attribut donné.

X=Attribut	x=Valeurs	P(Y=Jouer = Oui)=9/14	P(Y=Jouer = Non)=5/14
Temps	Ensoleillé	P(Ensoleillé/jouer=oui)=2/9	P(Ensoleillé/jouer)non)=3/5
	Couvert	P(Couvert//jouer=oui)=4/9	P(Couvert//jouer=non)=0/5
	Pluie	P(Pluie/jouer=oui)=3/9	P(Pluie/jouer=non)=2/5

Avec l'Estimateur de Laplace

X=Attribut	x=Valeurs	P(Y=Jouer = Oui)=9/14	P(Y=Jouer = Non)=5/14
Temps	Ensoleillé	P(Ensoleillé/jouer=oui)=2/9	P(Ensoleillé/jouer)non)=(3+1/(5+3)=4/8
	Couvert	P(Couvert//jouer=oui)=4/9	P(Couvert//jouer=non)=(0+1)/(5+3)=1/8
	Pluie	P(Pluie/jouer=oui)=3/9	P(Pluie/jouer=non)=(2+1)/(5+3)=3/8

L’Estimateur de Laplace

X=Attribut	x=Valeurs	P(Y=Jouer = Oui)=9/14	P(Y=Jouer = Non)=5/14
Temps	Ensoleillé	P(Ensoleillé/jouer=oui)=2/9	P(Ensoleillé/jouer)non)=4/8
	Couvert	P(Couvert//jouer=oui)=4/9	P(Couvert//jouer=non)=1/8
	Pluie	P(Pluie/jouer=oui)=3/9	P(Pluie/jouer=non)=3/8
Température	Chaude	P(Chaude/jouer=oui)=2/9	P(Chaude/jouer=non)=2/5
	Tiède	P(Tiède/jouer=oui)=4/9	P(Tiède/jouer=non)=2/5
	Fraiche	P(fraiche/jouer=oui)=3/9	P(Fraiche/jouer=non)=1/5
Humidité	Elevée	P(Elevée/jouer=oui)=3/9	P(Elevée/jouer=non)=4/5
	Normale	P(Normale/jouer=oui)=6/9	P(Normale/jouer=non)=1/5
Vent	Fort	P(Fort/jouer=oui)=3/9	P(Fort/jouer=non)=3/5
	Faible	P(Faible/jouer=oui)=6/9	P(Faible/jouer)non)=2/5
Jouer		9	5

Estimateur de Laplace

x = (Couvert, fraiche, *élevée*, *fort*)
On voudrait prédire la classe de la donnée X =*x*

$$\boldsymbol{Y_{MAP} = argmax \ \{p(Y = oui/X = x)\,, p(Y = non/X = x)\}}$$

$\boldsymbol{p(Y = oui\ /X = x)}$= $\boldsymbol{p(X = x/Y = oui\ ) \times p(Y = oui)}$

$\boldsymbol{p\big(Y = oui\ /X = (}couvert, fraiche, élevée, fort\boldsymbol{)\big)}$ = $\boldsymbol{p(X = (}couvert, fraiche, élevée, fort\boldsymbol{)/Y = oui\ ) \times}$
$\boldsymbol{p(Y = oui)}$

$\boldsymbol{p\big(Y = oui\ /X = (}couvert, fraiche, élevée, fort\boldsymbol{)\big)}$ = $\boldsymbol{p(X = }\ couvert\boldsymbol{/Y = oui\ ) \times}$
$\boldsymbol{p(X = }\ fraiche\boldsymbol{/Y = oui\ ) \times}$
$\boldsymbol{p(X = }\ élevée\boldsymbol{/Y = oui\ ) \times}$
$\boldsymbol{p(X = }\ fort\boldsymbol{/Y = oui\ )}$

$\boldsymbol{p\big(Y = oui\ /X = (}couvert, fraiche, élevée, fort\boldsymbol{)\big)}$ = $\frac{4}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} = \frac{4}{243}$

$\boldsymbol{p(Y = oui)}$= $\frac{9}{14}$

Estimateur de Laplace

x = (*couvert*, fraiche, *élevée*, *fort*)
On voudrait prédire la classe de la donnée X =*x*

$$Y_{MAP} = argmax \ \{p(Y = oui/X = x), p(Y = non/X = x)\}$$

$$p(Y = non\,/X = x) = p(X = x/Y = non) \times p(Y = non)$$

$$p\big(Y = non\,/X = (couvert, fraiche, élevée, fort)\big) = p(X = (couvert, fraiche, élevée, fort)/Y = non) \times p(Y = non)$$

$$p\big(Y = non/X = (ensoleillé, fraiche, élevée, fort)\big) = p(X = couvert/Y = non) \times p(X = fraiche/Y = non) \times p(X = élevée/Y = non) \times p(X = fort/Y = non)$$

$$p\big(Y = non/X = (couvert, fraiche, élevée, fort)\big) = \frac{1}{8} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} = \frac{36}{625}$$

$$p(Y = non) = \frac{5}{14}$$

Cas de Données manquante

x = (?, fraiche, *élevée*, *fort*)

On voudrait prédire la classe de la donnée X =*x*

$$p(Y = oui\ /X) = \frac{p(X/Y{=}oui\ ) \times p(Y{=}oui)}{p(X)} \quad et \quad p(Y = non\ /X) = \frac{p(X/Y{=}non\ ) \times p(Y{=}non)}{p(X)}$$

$$p(Y = oui/X = (?, \text{fraiche}, \textit{élevée}, \textit{fort})) = p(X = (?, \text{fraiche}, \textit{élevée}, \textit{fort})/Y = oui\ ) \times p(Y = oui)$$

$$p(Y = oui/X = (?, \text{fraiche}, \textit{élevée}, \textit{fort})) = p(temps = ?/Y = oui\ ) \times p(\text{temperature}{=}\text{fraiche}/Y = oui\ ) \times p(\text{humidité}{=}\textit{elevée}/Y = oui\ ) \times p(\text{vent}{=}\textit{fort}/Y = oui\ ) \times p(Y = oui)$$

$$p(Y = oui/X = (?, \text{fraiche}, \textit{élevée}, \textit{fort})) = 1 \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$$

Cas de Données manquante

x = (?, fraiche, *élevée*, *fort*)

On voudrait prédire la classe de la donnée X =*x*

$$p(Y = oui\ /X) = \frac{p(X/Y=oui\ ) \times p(Y=oui)}{p(X)} \quad et \quad p(Y = non\ /X) = \frac{p(X/Y=non\ ) \times p(Y=non)}{p(X)}$$

$p(Y = non/X =$ (?, fraiche, *élevée*, *fort*)$)$ = $p(X =$ (?, fraiche, *élevée*, *fort*)$/Y = non\ ) \times p(Y = non)$

$p(Y = non/X =$ (?, fraiche, *élevée*, *fort*)$)$ = $p(temps = ?/Y = non\ ) \times$
$p(\text{temperature=fraiche}/Y = non\ ) \times$
$p(\text{humidité=}elevée/Y = non\ ) \times$
$p(\text{vent=}fort/Y = non\ ) \times p(Y = non)$

$p(Y = non/X =$ (?, fraiche, *élevée*, *fort*)$)$ = $1 \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$

Cas d'Attribut numérique

Attribut numériques

Si l'attribut est numérique, on suppose que la distribution de la valeur de l'attribut $\boldsymbol{x_d}$ est normale moyenne μ_d et l'écart-type σ_d : $N(\mu_d, \sigma_d)$

$$p(\mathrm{x}_d = x_d | \mu_d, \sigma_d) = \frac{1}{\sigma_d \sqrt{2\pi}} \exp\left(-\frac{(x_d - \mu_d)^2}{2{\sigma_d}^2} \right)$$

Attribut numériques

Exemple

Jour	Temps	Température	humidité	vent	Jouer au tennis ?
1	Ensoleillé	27, 5	Élevée	Faible	Non
2	Ensoleillé	25	Élevée	Fort	Non
3	Couvert	26, 5	Élevée	Faible	Oui
4	Pluie	20	Élevée	Faible	Oui
5	Pluie	19	Normale	Faible	Oui
6	Pluie	17, 5	Normale	Fort	Non
7	Couvert	17	Normale	Fort	Oui
8	Ensoleillé	21	Élevée	Faible	Non
9	Ensoleillé	19, 5	Normale	Faible	Oui
10	Pluie	22, 5	Normale	Faible	Oui
11	Ensoleillé	22, 5	Normale	Fort	Oui
12	Couvert	21	Élevée	Fort	Oui
13	Couvert	25, 5	Normale	Faible	Oui
14	Pluie	20, 5	Élevée	Fort	Non

	Température	
	oui	non
	26, 5	27, 5
	20	25
	19	17, 5
	17	21
	19, 5	20, 5
	22, 5	
	22, 5	
	21	
	25, 5	
moyenne	21, 5	22, 3
écart-type	2, 91	3, 53

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Classificateur Naif Bayes

Exemple

x = (pluie, 18, *élevée*, *fort*)

$$p(Y = oui\ /X) = \frac{p(X/Y{=}oui\ ) \times p(Y{=}oui)}{p(X)}$$

$$(Y = non\ /X) = \frac{p(X/Y{=}non\ ) \times p(Y{=}non)}{p(X)}$$

	Température	
	oui	non
	26,5	27,5
	20	25
	19	17,5
	17	21
	19,5	20,5
	22,5	
	22,5	
	21	
	25,5	
moyenne	21,5	22,3
écart-type	2,91	3,53

$p(Y = oui/X =$ (pluie, 18, *élevée*, *fort*)$)$ = $p(X =$ (pluie, 18, *élevée*, *fort*)$/Y = oui\ ) \times p(Y = oui)$

$p(Y = oui/X =$ (pluie, 18, *élevée*, *fort*)$)$ = $p(temps = \text{pluie}/Y = oui\ ) \times p(\text{temperature}{=}18/Y = oui\ ) \times p(\text{humidité}{=}elevée/Y = oui\ ) \times p(\text{vent}{=}fort/Y = oui\ ) \times p(Y = oui)$

$$p(\text{temperature}{=}18/Y = oui\ ) = \frac{1}{2.91 * \sqrt{2\pi}} e^{-\frac{(18-21.5)^2}{2*(21.5)^2}}$$

$p(Y = oui/X =$ (pluie, 18, *élevée*, *fort*)$)$ $= \frac{3}{9} \times 0{,}066 \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}$

Exemple

x = (pluie, 18, *élevée*, *fort*)

$$p(Y = oui\ /X) = \frac{p(X/Y{=}oui) \times p(Y{=}oui)}{p(X)}$$

$$(Y = non\ /X) = \frac{p(X/Y{=}non) \times p(Y{=}non)}{p(X)}$$

	Température	
	oui	non
	26,5	27,5
	20	25
	19	17,5
	17	21
	19,5	20,5
	22,5	
	22,5	
	21	
	25,5	
moyenne	21,5	22,3
écart-type	2,91	3,53

$p(Y = non/X =$ (pluie, 18, *élevée*, *fort*)) = $p(X =$ (pluie, 18, *élevée*, *fort*)$/Y = non) \times p(Y = non)$

$p(Y = non/X =$ (pluie, 18, *élevée*, *fort*)) = $p(temps = \text{pluie}/Y = non) \times$
$p(\text{temperature}{=}18/Y = non) \times$
$p(\text{humidité}{=}elevée/Y = non) \times$
$p(\text{vent}{=}fort/Y = non) \times p(Y = non)$

$$p(\text{temperature}{=}18/Y = non) = \frac{1}{3.53 * \sqrt{2\pi}}\, e^{-\frac{(18-22.3)^2}{2*(22.3)^2}}$$

$p(Y = non/X =$ (pluie, 18, *élevée*, *fort*)) $= \frac{3}{5} \times \mathbf{0,053} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}$

Evaluation

VP : le nombre de données de la classe positive qui ont correctement classées dans la classe positive

VN : : le nombre de données de la classe négative qui ont correctement classées dans la classe négative

$\mathrm{F}P$: le nombre de données de la classe négative qui ont classées dans la classe positive

$\mathrm{F}N$: le nombre de données de la classe positive qui ont classées dans la classe négative

Evaluation

La matrice de confusion

Etiquettes Prédites

Etiquettes réelles		+	-
	+	VP	FN
	-	FP	VN

Evaluation

Taux de bonne classification (accuracy)

C'est le nombre de données qui ont été mal classés

$$\frac{VP + VN}{VP + VN + FP + FN}$$

Erreur de classification= 1-accuracy

Classificateur Naif Bayes

Evaluation

- **précision:** le nombre de données correctement classées dans la classe positive (resp. négative) sur le nombre total de données prédits positives (resp. négatives).

-***précision*** *(classe positive)*= $\frac{VP}{VP+FP}$

- ***précision****(classe négative)* =$\frac{VN}{VN+FN}$

Classificateur Naif Bayes

Evaluation

- **rappel :** le nombre de données correctement classées positives (resp. négative) sur le nombre total de données réellement positives (resp. négative)

***Rappel** (classe positive)* $= \frac{VP}{VP+FN}$

***Rappel** (classe négative)* $= \frac{VN}{FP+VN}$