LOG6308 — Systèmes de recommandation

Rappel de la théorie de Bayes

Michel C. Desmarais

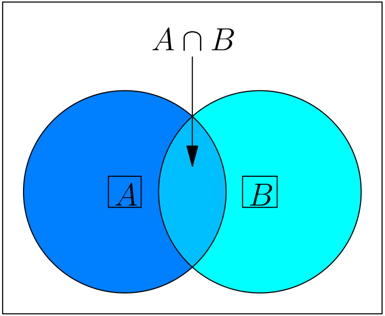
Génie informatique et génie logiciel
École Polytechnique de Montréal

Automne, 2017
(version 13 septembre 2017)

Rappel de la théorie de Bayes

1. Probabilités conditionnelles et règle de Bayes
2. Version du ratio de chances (odds-likelihood)
3. Estimation par la vraisemblance maximale (maximum-likelihood)

La perspective ensembliste de la probabilité conditionnelle



Dérivation de la règle de Bayes :

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A|B)P(B) &= P(A \cap B) \\
P(A \cap B) &= P(B|A)P(A) \\
P(A|B)P(B) &= P(B|A)P(A) \\
P(A|B) &= \frac{P(B|A)P(A)}{P(B)}
\end{aligned}
$$

Fréquences et probabilités conditionnelles

A	B	Fréq
1	1	x_1
1	0	x_2
0	1	x_3
0	0	x_4

$$
\begin{aligned}
P(A=1|B=1) &= \frac{f(A=1,B=1)}{f(B)} \\
&= \frac{f(A=1,B=1)}{f(B=1,A=1)+f(B=1,A=0)} \\
&= \frac{f(x_1)}{f(x_1)+f(x_3)}
\end{aligned}
$$

Dans une forme plus générale, on obtient :

$$
P(A|B) = \frac{P(A,B)}{\sum_i P(B|A_i)P(A_i)}
$$

Notion de probabilité conditionnelle

La probabilité conditionnelle peut être vue comme une probabilité mise dans un contexte plus spécifique.

La Règle de Bayes permet d'obtenir une probabilité conditionnelle, $P(A|B)$, sans pour autant avoir les données sur la fréquence $P(A, B)$, mais plutôt à partir des fréquences de $P(B|A)$ et $P(A)$.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Problème 1

Deux compagnies de taxi opèrent à Belleville, les bleus et les verts, respectivement de Taxis Belleville Bleus et Taxis Belleville Verts. 85% des taxis sont des bleus et 15% sont des verts. Un taxi frappe une auto stationnée la nuit et s'enfuit. Un témoin identifie la couleur du taxi comme étant verte. Un procès s'en suit et des tests permettent d'établir qu'en moyenne seulement 80% des gens réussissent à identifier correctement la couleur d'un véhicule la nuit. Quelle est la probabilité que le témoin ait bien identifié la couleur du véhicule ?

Règle de Bayes

La formulation classique de la règle, ou théorème de Bayes :

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

où :

$P(A)$ est la probabilité initiale de A (*prior probability*) ;
$P(B|A)$ est la vraisemblance de B étant donné A ;
$P(B)$ est la probabilité initiale de B

ÉCOLE POLYTECHNIQUE MONTRÉAL

Solution au problème 1

Posons :

$P(Ta{=}V) = 0,15$: la probabilité qu'un taxi soit *vert*.

$P(Ta{=}B) = 0,85$: la probabilité qu'un taxi soit *bleu*.

$P(Te{=}V \mid Ta{=}V) = 0,80$: la probabilité qu'un témoin identifie un taxi *vert* la nuit étant donné que le taxi est *vert*.

$P(Te{=}V \mid Ta{=}B) = 0,20$: la probabilité qu'un témoin identifie un taxi *vert* la nuit étant donné que le taxi est *bleu*.

On cherche $P(Ta{=}V \mid Te{=}V)$.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Solution au problème 1 (suite)

Selon la règle de Bayes :

$$
\begin{aligned}
P(Ta{=}V \mid Te{=}V) & \\
&= \\
&= \frac{P(Te{=}V \mid Ta{=}V)P(Ta{=}V)}{P(Te{=}V)} \\
&= \frac{0,8 \times 0,15}{\color{red}{P(Te{=}V)}} \\
&= \frac{0,8 \times 0,15}{P(Te{=}V \mid Ta{=}V)P(Ta{=}V) + P(Te{=}V \mid Ta{=}B)P(Ta{=}B)} \\
&= \frac{0,8 \times 0,15}{(0,8 \times 0,15) + (0,2 \times 0,85)} \\
&= \frac{0,8 \times 0,15}{(0,8 \times 0,15) + (0,2 \times 0,85)} \\
&= 0,414
\end{aligned}
$$

ÉCOLE POLYTECHNIQUE MONTRÉAL

Factorisation en probabilités conditionnelles

Par définition :

$$P(A) = P(A,B) + P(A,\overline{B})$$

or, selon la règle de Bayes,

$$\begin{aligned} P(A,B) &= P(A|B)P(B) \\ P(A,\overline{B}) &= P(A|\overline{B})P(\overline{B}) \end{aligned}$$

par substitution, on obtient

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

Exercice

Supposons qu'un test pour le virus VIH donne un résultat positif à 90% lorsqu'un individu est effectivement infecté, et qu'il donne un (faux) résultat positif à 20% lorsqu'un individu n'est pas infecté. Sachant que 0,1% de la population est infectée, quelle est la probabilité qu'un individu qui subit le test soit infecté si le résultat est positif ?

Les chances et les probabilités

Comme la somme des probabilités est de 1,

$$P(A) + P(\overline{A}) = 1 \Rightarrow P(A) = 1 - P(\overline{A})$$

Les chances de A sont, par définition,

$$O(A) = \frac{P(A)}{P(\overline{A})} = \frac{P(A)}{1 - P(A)}$$

Donc : $$P(A) = \frac{O(A)}{1 + O(A)}$$

De même pour $O(H|E)$: $$O(H|E) = \frac{P(H|E)}{P(\overline{H}|E)}$$

et ainsi : $$P(H|E) = \frac{O(H|E)}{1 + O(H|E)}$$

Rappel de la théorie de Bayes

ÉCOLE POLYTECHNIQUE MONTRÉAL

Version du ratio de chances (*odds-likelihood*)

Lorsque les événements peuvent prendre deux valeurs, notamment *vrai* ou *faux*, la version du ratio de chances s'avère particulièrement utile. Elle évite le calcul de $P(e)$, la probabilité initiale de l'évidence, c.-à-d. le dénominateur. Posons H une hypothèse à l'étude et e un nouvel indice (*évidence*). Alors

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(H)}{P(\overline{H})}\frac{P(E|H)}{P(E|\overline{H})}$$

Cette expression fournit les *chances postérieures* (*posterior odds*) des hypothèses. Les chances initiales sont $\frac{P(H)}{P(\overline{H})}$ et le ratio de vraisemblance est $\frac{P(E|H)}{P(E|\overline{H})}$.

Retour sur le VIH

Posons : E : test positif ($E = 1$)
H : patient infecté ($H = 1$)

On a les probabilités suivantes :

$P(E|H) = 0,9$ Un patient infecté a un résultat positif

$P(E|\overline{H}) = 0,2$ Un patient non infecté a un résultat positif

$P(H) = 0,001$ Un patient est infecté

$P(\overline{H}) = 0,999$ Un patient n'est pas infecté

On peut alors faire le calcul suivant :

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(H)}{P(\overline{H})}\frac{P(E|H)}{P(E|\overline{H})} = \frac{0,001}{0,999} \times \frac{0,9}{0,2} = 0,0045$$

Hypothèse d'inépendance des évidences

Si on présume l'indépendance suivante :

$$P(E_1, E_2|H) = P(E_1|H)P(E_2|H)$$

il devient alors possible de combiner les évidences par le produit des ratios de chances :

$$\frac{P(H|E_1, E_2)}{P(\overline{H}|E_1, E_2)} = \frac{P(H)}{P(\overline{H})}\frac{P(E_1|H)}{P(E_1|\overline{H})}\frac{P(E_2|H)}{P(E_2|\overline{H})}$$

La combinaison des évidences

De manière plus générale, on peut définir deux types de vraisemblances. La *vraisemblance de suffisance*, LS_{EH}, et la *vraisemblance de nécessité*, LN_{EH} :

$$LS_{EH} = \frac{P(E \mid H)}{P(E \mid \overline{H})} \qquad LN_{EH} = \frac{P(\overline{E} \mid H)}{P(\overline{E} \mid \overline{H})}$$

Ces définitions nous permettent d'obtenir les ratios de chances :

$$\begin{aligned} O(H \mid E) &= LS_{EH}\, O(H) \\ O(H \mid \overline{E}) &= LN_{EH}\, O(H) \end{aligned}$$

De là, sous l'hypothèse d'indépendance mentionnée, on obtient la formulation générale :

$$O(H \mid E_1, E_2, \ldots, E_n) = O(H) \prod_i^n L\bullet_{E_i H}$$

où

$L\bullet_{E_i H}$ représente $LS_{E_i H}$ ou $LN_{E_i H}$ selon le cas.

Exercice

Quelle est la probabilité qu'un patient soit infecté au VIH s'il teste positif trois fois ?

Quelle est l'hypothèse qu'il faut faire ?

En présumant qu'il existe un médicament qui, losrqu'ingurgité, diminue la discrimination du test, est-ce une violation de l'hypothèse en question ?

Rappel de la théorie de Bayes

1. Probabilités conditionnelles et règle de Bayes
2. Version du ratio de chances (odds-likelihood)
3. Estimation par la vraisemblance maximale (maximum-likelihood)

Estimation de paramètres avec la vraisemblance maximale

La vraisemblance maximale, ou *maximum likelihood*, est une méthode très répandue pour estimer des paramètres que l'on observe pas directement. On s'en sert notamment pour l'entraînement du modèle bayésien naïfs (*Naive Bayes*).

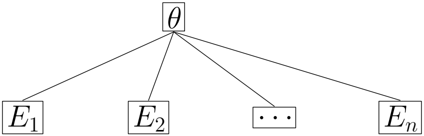
Le principe consiste à définir l'influence d'une variable (paramètre), θ, sur un ensemble de variables observées, $E_1, E_2, \cdots, E_n$, par un modèle probabiliste. Puis, en faisant une hypothèse d'indépendance entre les observations, on calcule la probabilité d'observer une configuration donnée en fonction de la valeur de θ.

Il s'agit finalement de trouver la valeur de θ la plus vraisemblable étant donné les observations $E_1{=}e_1, E_2{=}e_2, \cdots, E_n{=}e_n$.

ÉCOLE POLYTECHNIQUE MONTRÉAL

Probabilité d'observations multiples sous l'hypothèse d'indépendance

Prenons un paramètre, θ, qui influence un certain nombre de paramètres, $E_1, E_2, \cdots, E_n$.



En présumant l'indépendance entre deux évidences étant donné le paramètre θ, la probabilité d'observer une série d'observations données, $E_1{=}e_1, E_2{=}e_2, \cdots, E_n{=}e_n$ est :

$$P(e_1, e_2, \cdots, e_n|\theta) = kP(\theta)\prod_i P(e_i|\theta)$$

Estimation avec la log-vraisemblance maximale (*maximum log-likelihood*)

L'équation :

$$P(e_1, e_2, \cdots, e_n|\theta) = kP(\theta)\prod_i P(e_i|\theta)$$

peut se réécrire sous la forme logarithmique :

$$\log(P(e_1, e_2, \cdots, e_n|\theta)) = \sum_i \log(P(e_i|\theta))$$

Cette dernière formulation facilite l'application de techniques d'optimisation standard pour trouver un maximum. Par exemple, lorsque la dérivée première existe, on peut trouver le maximum de manière analytique.

ÉCOLE POLYTECHNIQUE MONTRÉAL

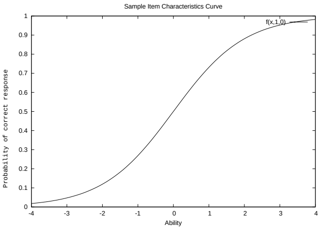
Un exemple

Présumons que le succès d'un individu à répondre correctement à une série de questions, $(X_1, X_2, \ldots, X_n)$, dépend de son habileté, θ. La probabilité d'observer une série spécifique de réponses, $(x_1, x_2, \ldots, x_n)$ est donc :

$$P(x_1, x_2, \cdots, x_n|\theta) = kP(\theta)\prod_i P(x_i|\theta)$$

Le problème consiste donc à trouver la valeur de θ qui maximisera cette probabilité. Il nous faut un modèle de $P(X_i|\theta)$ comme celui de la figure ci-contre, où :

$$P(X_i|\theta) = \frac{1}{1+e^{-\theta}}$$



Un exemple (suite)

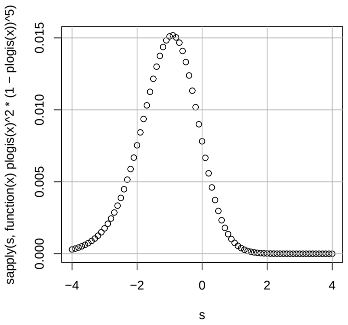
Si notre individu a répondu à 7 questions et obtenu :

$$(0, 1, 1, 0, 0, 0, 0)$$

alors la probabilité d'observer cette séquence s'il avait une habileté $\theta = 1$ serait proportionnelle à

$$0.731^2 \times (1 - 0.731)^5 = 0,001$$

car
$P(X_i{=}1|\theta{=}1) = \frac{1}{1+e^{-1}} = 0,731.$
La probabilité pour $\theta{=}-1$, elle, correspond à $0,015$



Le graphique ci-dessus indique la probabilité pour les différentes valeurs de θ. La valeur maximale se situe à environ $\theta = -0,92$