Apprentissage statistique I

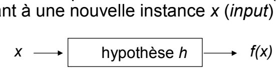
IFT-17587
Concepts avancés pour systèmes intelligents
Luc Lamontagne

Plan

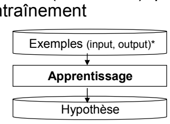
- Approche probabiliste - *Naive Bayes*
- Apprentissage à base d'exemples (*Instance-based*)
- Réseaux de neurones

Rappel sur l'apprentissage inductif

- Classification
  - Utiliser l'hypothèse *h* pour déterminer la valeur (*output*) correspondant à une nouvelle instance *x* (*input*)

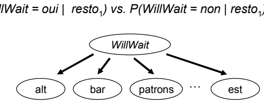


- Apprentissage
  - Construire l'hypothèse *h* (un modèle) qui recouvre le mieux nos exemples d'entraînement



Approche probabiliste

- Problème de classification
  - Sélectionner la valeur C qui maximise $P(C \mid x_1, \ldots x_n)$
    - $x_1, \ldots x_n$ sont les valeurs des attributs de l'exemple à classifier
    - Exemple du restaurant :
      - Décision à prendre pour un restaurant *resto*$_1$
        resto$_1$: *alt = yes, bar = yes, …, est = 0-10*
      - La décision consiste à retenir la probabilité maximum entre
        P(WillWait = oui | resto$_1$*) vs. P(WillWait = non | resto*$_1$*)*



  - Il nous faut toutefois le modèle probabiliste *P* → hypothèse…

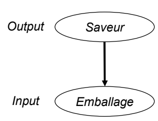
Approche probabiliste :

Apprentissage du modèle

- *Règle de Bayes*

$$P(C \mid x_1, \ldots x_n) = \alpha\, P(x_1, \ldots x_n \mid C)\, P(C)$$

- Apprentissage du modèle
  - Trouver *la meilleure hypothèse h pour nos exemples*
    - On doit donc déterminer les distributions $P(x_1, \ldots x_n \mid C)$ et $P(C)$



$P(C = cerise) = \theta$

Saveur	P(Emballage = rouge| Saveur)
Cerise	θ_1
Lime	θ_2

Éléments de l'hypothèse h

Approche probabiliste :

Exemple des sacs de bonbons

- Un manufacturier produit des bonbons en deux saveurs différentes :
  - Cerise et lime
- Chacun bonbon est enveloppé dans un emballage opaque
  - On ne peut pas connaître la saveur du bonbon sans retirer l'emballage.
- Le manufacturier offre 5 sortes de sacs préparés à l'avance dont les proportions varient :
  - h_1 : 100% cerise
  - h_2 : 75% cerise + 25% lime
  - h_3 : 50% cerise + 50% lime
  - h_4 : 25% cerise + 75% lime
  - h_5 : 100% lime
- Tous les sacs sont identiques de l'extérieur !
  - Le seul moyen de distinguer les sacs est de piger des bonbons à l'intérieur du sac et de les développer pour connaître la saveur.

Approche probabiliste – Exemple des bonbons :

Version simple de l'exemple

- Supposons :
  - On vous remet un sac .
  - Vous ne connaissez pas sa distribution de bonbons.
  - 5 hypothèses possibles : h_1 , h_2 , h_3 , h_4 , h_5 .
- Comment faire pour déterminer le type de sac?
  - On pige un nombre suffisamment élevé de bonbons.
    - Les bonbons b_i que l'on pige sont nos exemples. …
  - A partir de ces exemples :
    - On estime la probabilité que le sac soit de type h_i → $P(h_i \mid b_1, b_2 \ldots b_N)$
    - Par exemple : $P($ | … $)$
  - On retient le type de sac qui est le plus probable.
    - On retient l'hypothèse qui explique le mieux la distribution de nos exemples.
    - Choisir la meilleure hypothèse → apprentissage

Approche probabiliste – Exemple des bonbons :

Version simple de l'exemple

- Par la règle de Bayes

$$P(h_i \mid b_1, b_2 \ldots b_N) = \alpha\, P(b_1, b_2 \ldots b_N \mid h_i)\, P(h_i)$$

- $P(h_i)$ *: la distribution à priori des sacs*
  - On suppose que c'est déterminé par le fabriquant
  - Par exemple la distribution suivante :

- $P(b_1, b_2 \ldots b_N \mid h_i)$ *: la vraisemblance de piger N bonbons*
  - On suppose que chaque pige est indépendante des autres.

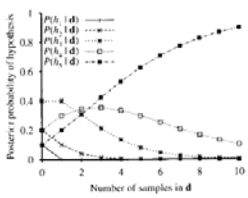
$$P(b_1, \ldots, b_N \mid h_i) = \prod_j P(b_j \mid h_i)$$

  - Exemple:
    - 10 bonbons pigés :
    - $P($ | $) = (0.5)^{10} = 0.000244$
    - $P($ | $) = \alpha \,.\, (0.000244) \,.\, (0.4) = 0.000097\, \alpha$

Approche probabiliste – Exemple des bonbons :

Version simple de l'exemple

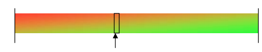
- Comparaison des hypothèses
  - Exemples b_i :
  - Estimation de la probabilité à posteriori $P(h_i \mid b_1, b_2 .... b_N)$



Approche probabiliste – Exemple des bonbons :

Version un peu moins simple

- Maintenant il n'y a plus de type de sac préparé à l'avance…
  - La proportion cerise/lime varie d'un sac à l'autre
  - Le % de cerise = θ (% de lime : $1 - \theta$)
  - Les distributions de sac sont équiprobables



- $P(b_1, b_2 .... b_N \mid h_\theta)$: *la probabilité de piger N bonbons pour l'hypothèse θ*
  - Si N contient c et l
  - La vraisemblance de cet échantillon d'exemple est

$$P(b_1,\ldots,b_N \mid h_\theta) = \prod_j P(b_j \mid h_\theta) = \theta^c.(1-\theta)^l$$

Approche probabiliste – Exemple des bonbons :

Version un peu moins simple

- Quel est la meilleure hypothèse h_θ?
  - Celle qui recouvre le mieux les N exemples (c et l)
- Donc trouver le θ qui maximise la vraisemblance $P(b_1, b_2 .... b_N \mid h_\theta)$
  - Plus facile de maximiser $\log P(b_1, b_2 .... b_N \mid h_\theta)$

$$\log P(b_1,\ldots,b_N \mid h_\theta) = \log\left(\theta^c.(1-\theta)^l\right) = c\log\theta + l\log(1-\theta)$$

  - Pour maximiser : on fait la dérivé et on rend l'expression égale à 0

$$\frac{\partial \log P(b_1,\ldots,b_N \mid h_\theta)}{\partial\theta} = \frac{c}{\theta} - \frac{l}{(1-\theta)} = 0$$

$$\theta = \frac{c}{c+l} = \frac{c}{N}$$ ⇦ C'est tout simplement la proportion de !

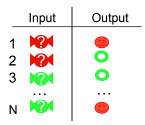
Apprentissage probabiliste :

Maximum de vraisemblance

- Meilleure hypothèse
  - Hypothèse contenant des paramètres continus (probabilités)
  - Trouver les paramètres qui optimise $P(d|h)$ pour nos exemples d'entraînement
- Approche du maximum de vraisemblance
  - Formuler l'expression $P(d|h)$
  - On prend le logarithme et on dérive
  - On trouve les paramètres lorsque la dérivé est égale à 0

Approche probabiliste – Exemple des bonbons :
Version encore un peu moins simple

- Chacun des N bonbons est maintenant enveloppé dans un emballage rouge ou vert
  - *Input :*
    - r_c *cerises en rouge*
    - g_c *cerises en vert*
    - r_l *limes rouge*
    - g_l *cerises vert*
  - *Output :*
    - c *cerises +* l *limes*

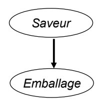


- *Paramètres du modèle à estimer*

$$P(C = cerise) = \theta$$

$$P(Emballage = rouge \mid cerise) = \theta_1$$

$$P(Emballage = rouge \mid lime) = \theta_2$$



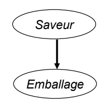
Approche probabiliste – Exemple des bonbons :
Version encore un peu moins simple

- *Probabilité de nos exemples d'entraînement* d

$$P(d \mid h_{\theta,\theta_1,\theta_2}) = \prod_j P(d_j \mid h_{\theta,\theta_1,\theta_2}) = \theta^c.(1-\theta)^l.\theta_1^{r_c}.(1-\theta_1)^{g_c}.\theta_2^{r_l}.(1-\theta_2)^{g_l}$$

- *Maximum de vraisemblance* *(les résultats seulement…)*

$P(C = cerise) = \theta \;=\; c/(c+l)$ ← *Proportion de cerises dans la colonne output*



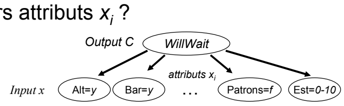
Saveur	$P(\text{Emballage} = rouge \mid Saveur)$
cerise	$\theta_1 = r_c / (r_c + g_c)$
lime	$\theta_2 = r_l / (r_l + g_l)$

← *Proportion de cerises qui sont emballées en rouge selon notre tableau*

Apprentissage : on compte le nombre d'occurences dans notre tableau d'exemples!

Apprentissage probabiliste :
Naïve Bayes

- Que faire si notre variable de classe *C* dépend de plusieurs attributs x_i ?



  - Approximation : les attributs sont mutuellement indépendants

$$P(C \mid x_1,\dots,x_n) = \alpha P(C)\prod_i P(x_i \mid C)$$ ← Naïve Bayes!

  - Pour notre exemple du restaurant

$$P(C = WillWait \mid x) = \alpha.P(WillWait).P(Alt = yes \mid WillWait).P(Bar = yes \mid WillWait)....$$

Apprentissage probabiliste :
Naïve Bayes

- Donne de bons résultats en pratique
- Particulièrement efficace avec une approche ensemble de type *boosting*
- Un problème :
  - Si $x_i{=}a$ et $C{=}b$ n'est pas dans nos exemples
  - Alors $P(C{=}b \mid x_i{=}a) = 0$
- Il existe différentes fonctions de lissage (*smoothing*) pour éliminer ce problème
  - Par exemple, additionner 1 à tous les comptes (*add-one*).