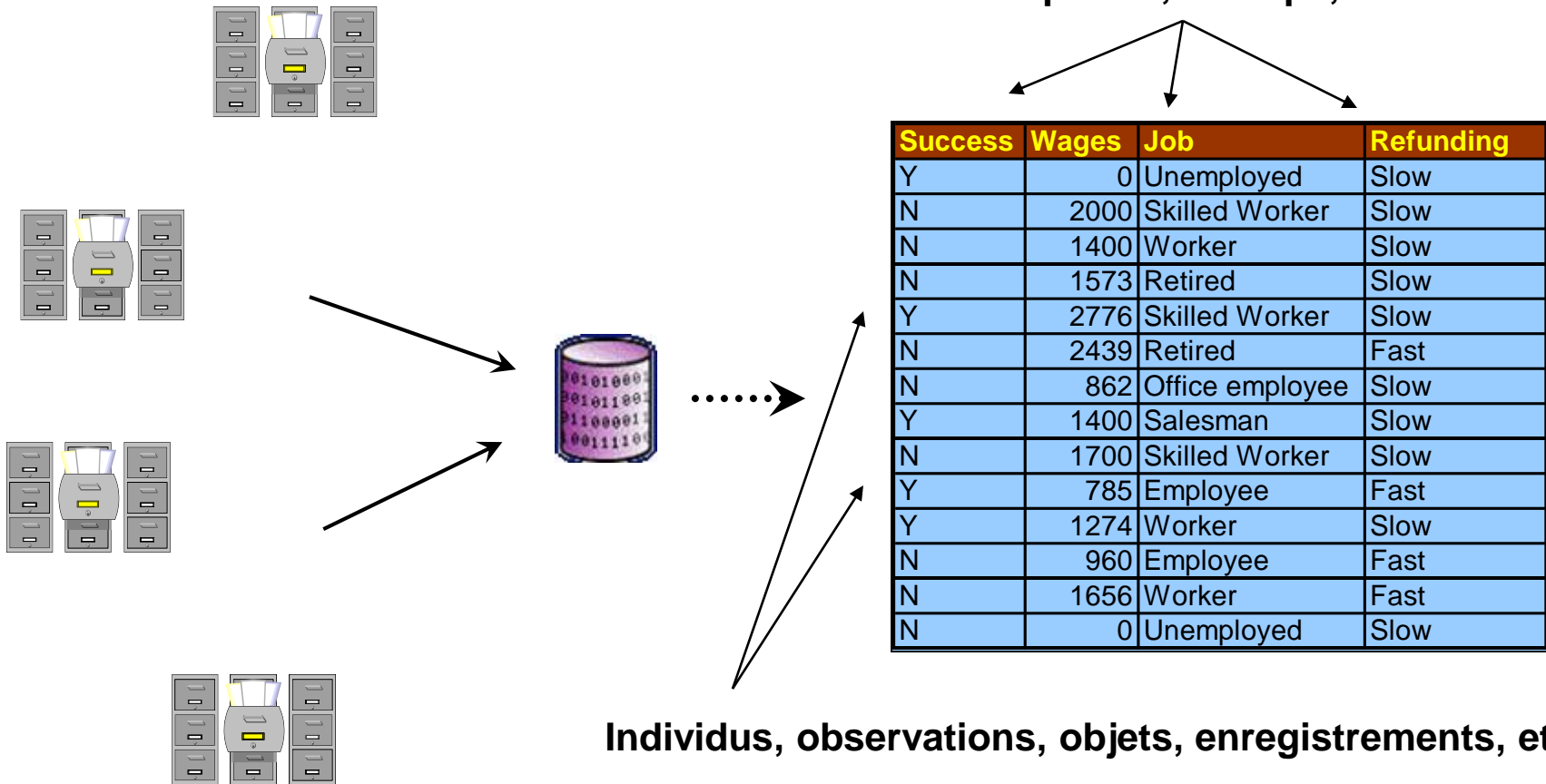


Apprentissage supervisé

Ricco Rakotomalala
Université Lumière Lyon 2

Tableau de données

Variables, caractères, attributs,
Descripteurs, champs, etc.



Individus, observations, objets, enregistrements, etc.

Statut des variables

Success	Wages	Job	Refunding
Y	0	Unemployed	Slow
N	2000	Skilled Worker	Slow
N	1400	Worker	Slow
N	1573	Retired	Slow
Y	2776	Skilled Worker	Slow
N	2439	Retired	Fast
N	862	Office employee	Slow
Y	1400	Salesman	Slow
N	1700	Skilled Worker	Slow
Y	785	Employee	Fast
Y	1274	Worker	Slow
N	960	Employee	Fast
N	1656	Worker	Fast
N	0	Unemployed	Slow

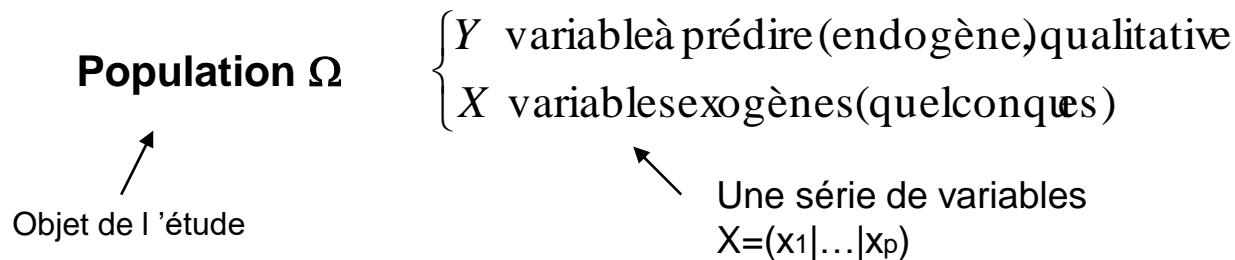
Variable à prédire
Attribut classe
Variable endogène

Nécessairement discrète nominale
(qualitative)

Variables prédictives
Descripteurs
Variables exogènes

De type quelconque
(nominale, ordinale, continue)

Principes de l'apprentissage supervisé



On veut construire une fonction de classement telle que

$$Y = f(X, \alpha)$$

Objectif de l'apprentissage

Utiliser un échantillon Ω_a (extraite de la population) pour choisir la fonction f et ses paramètres α telle que l'on minimise l'erreur théorique

$$ET = \frac{1}{\text{card}(\Omega)} \sum_{\Omega} \Delta[Y, \hat{f}(X, \hat{\alpha})]$$

$$\text{où } \Delta[.] = \begin{cases} 1 \text{ si } Y \neq \hat{f}(X, \hat{\alpha}) \\ 0 \text{ si } Y = \hat{f}(X, \hat{\alpha}) \end{cases}$$

Problèmes :

- \rightarrow il faut choisir une famille de fonction
- \rightarrow il faut estimer les paramètres α
- \rightarrow on utilise un échantillon pour optimiser sur la population

Apprentissage bayésien

(cas particulier du problème à 2 classes - Positifs vs. Négatifs)

Apprentissage en 2 étapes à partir des données :

- estimer la probabilité d'affectation $P(Y / X)$
- prédire $[Y = +]$ si $P(Y = + / X) > P(Y = - / X)$

Remarques :

- $P(Y = + / X)$ est selon le cas appelé « score » ou « appétence » : c'est la « propension à être un positif »
- Cette méthode d'affectation minimise l'erreur de prédiction -- c'est un cas particulier du coût de mauvaise affectation

Apprentissage bayésien

(généralisation à K classes)

Apprentissage en 2 étapes à partir des données :

- estimer la probabilité d'affectation $P(Y = y_k / X)$
- prédire $y_{k^*} = \arg \max_k P(Y = y_k / X)$

Remarque : Lorsque les X sont discrets, nous pouvons en déduire un modèle logique d'affectation.

Si $X_1 = ?$ et $X_2 = ?$ et $X_3 = ?$... Alors $Y = ?$

prémisse

conclusion

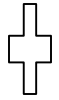
Apprentissage bayésien -- Exemple

Y X

Maladie	Poids	Taille	Marié	Etud.Sup
Présent	45	Trapu	Non	Oui
Présent	57	Elancé	Non	Oui
Absent	59	Elancé	Non	Non
Absent	61	Trapu	Oui	Oui
Présent	65	Elancé	Non	Oui
Absent	68	Elancé	Non	Non
Absent	70	Trapu	Oui	Non
Présent	72	Trapu	Non	Oui
Absent	78	Trapu	Oui	Non
Présent	80	Elancé	Oui	Non

- SI taille = ? ALORS Maladie = ?
- SI taille = ? ET etud.sup = ? ALORS Maladie = ?

Avantages et inconvénient du modèle bayésien complet



Optimale, elle minimise l'erreur théorique



Pas de solution directe pour les descripteurs continus
(*discrétisation ou hypothèse de distribution*)



Pas de sélection et d'évaluation des descripteurs
(*individuellement ou des groupes de variables - donc pas de sélection*)

Dès que le nombre de descripteurs augmente

- *Problème de calculabilité*



Nombre d'opérations énorme, ex. 10 descr. Binaires => 2^{10} règles

- *Problème de fragmentation des données*

Plein de cases avec des 0, estimations peu fiables

Cette approche n'est pas utilisable dans la pratique !

Évaluation de l'apprentissage

Compréhensibilité

Le modèle exprime une « **connaissance** »

- ① **Explication** : comprendre la causalité pour mieux l'exploiter
- ① **Validation** : l'expert peut évaluer la pertinence de l'expertise
- ① **Amélioration** : l'expert peut intervenir pour ajuster les paramètres calculés (ex. les bornes de discrétisation)

Rapidité

① En **apprentissage** → pouvoir tester plusieurs pistes (ajout de variables, test de combinaison de variables, modifications de paramètres, etc.)

- ① En **classement**, affecter une étiquette à un nouvel individu
- ① Facilité de **mise à jour** du modèle (cf. la notion d'incrémentalité)



Précision

① Évaluer la **précision** (qualité) du modèle lors de son utilisation future

Évaluation de l'apprentissage – Matrice de confusion

Principe : confronter la valeur observée avec la prédiction

		Prédite		Total
		+	-	
Observée	+	a	b	a+b
	-	c	d	c+d
	Total	a+c	b+d	n

Quelques indicateurs :

- Vrais positifs $VP = a$
- Faux positifs $FP = c$
- Taux d'erreur = $(c+b)/n$
- Sensibilité = Rappel = Taux de VP = $a/(a+b)$
- Précision = $a/(a+c)$
- Taux de FP = $c/(c+d)$
- Spécificité = $d/(c+d) = 1 - \text{Taux de FP}$

Évaluation – Les coûts de mauvaise affectation

Comparaison de deux méthodes d'apprentissage

		Prédite		Total
		+	-	
Observée	+	40	10	50
	-	20	30	50
	Total	60	40	100

		Prédite		Total
		+	-	
Observée	+	20	30	50
	-	0	50	50
	Total	20	80	100

☞ *Calculer les indicateurs synthétiques et comparer*

Une information complémentaire

La matrice de coûts de mauvais classement

		Prédite	
		+	-
Observée	+	0	5
	-	1	0

☞ *Coût moyen de mauvaise affectation (dont le taux d'erreur est un cas particulier)*

Évaluation – Le principe apprentissage & test

Problème : un fichier ne peut pas être juge et partie

Dans ce cas, les indicateurs calculés sont dit « de resubstitution »

On sait qu'ils sont biaisés -- trop optimistes

Success	Wages	Job	Refunding
Y	0	Unemployed	Slow
N	2000	Skilled Worker	Slow
N	1400	Worker	Slow
N	1573	Retired	Slow
Y	2776	Skilled Worker	Slow
N	2439	Retired	Fast
N	862	Office employee	Slow
Y	1400	Salesman	Slow
N	1700	Skilled Worker	Slow
Y	785	Employee	Fast
Y	1274	Worker	Slow
N	960	Employee	Fast
N	1656	Worker	Fast
N	0	Unemployed	Slow

Subdivision aléatoire

Échantillon d'apprentissage
Utilisé pour la construction du modèle
70%

Échantillon test
Utilisé pour l'évaluation du modèle
30%

☞ Rappel, précision, taux d'erreur...

(exercice : fichier LOAN - Success vs. Housing & Refunding)...

Bibliographique : compréhension des méthodes supervisées

- « Analyse discriminante - Application au risque et au scoring financier », M. Bardos, ed. Dunod, 2001.

Technique pratique, avec de bons repères théoriques, tourné vers les applications

- « The elements of statistical learning - Data Mining, Inference and Prediction », T. Hastie, R. Tibshirani, J. Friedman, Springer 2001.

Très technique, encyclopédique, indispensable pour la recherche, à lire plusieurs fois