

- Apprentissage probabiliste : calcule explicitement les probabilités des hypothèses, une des approches les plus pragmatiques pour certains types d'apprentissage
- Incrémental : chaque exemple met à jour la probabilité qu'une hypothèse est correcte. Des connaissances a priori peuvent être combinées avec des données d'observation.
- Prédiction probabiliste : prédit plusieurs hypothèses, pondérées par leur probabilité

Le problème de classification peut être formulé en utilisant les probabilités *a posteriori* :

$P(C|X)$  = probabilité que l'objet  $X = \langle x_1, \dots, x_k \rangle$  est de la classe  $C$

Exemple :  $P(\text{Classe}=\text{P} \mid X=\langle \text{ciel}=\text{dégagé}, \text{température}=\text{chaud}, \text{humidité}=\text{forte}, \text{vent}=\text{faux} \rangle)$

Idée : attribuer à l'objet  $X$  la classe  $C$  qui maximise  $P(C|X)$

$P(\text{Classe}=\mathbf{P} \mid X=\langle \text{ciel}=\text{dégagé}, \text{température}=\text{chaud}, \text{humidité}=\text{forte}, \text{vent}=\text{faux} \rangle)$

ou

$P(\text{Classe}=\mathbf{N} \mid X=\langle \text{ciel}=\text{dégagé}, \text{température}=\text{chaud}, \text{humidité}=\text{forte}, \text{vent}=\text{faux} \rangle)$

Théorème de Bayes

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)}$$

$P(\text{classe} | \text{observations}) = P(\text{observations} | \text{classe}) \cdot P(\text{classe}) / P(\text{observations})$

Principe : attribuer la classe la plus probable

trouver la classe  $C_i$  telle que  $P(C_i|X)$  est maximum  $\rightarrow C_i$  telle que  $P(X|C_i) \cdot P(C_i) / P(X)$  est maximum

**remarque** :  $P(X)$  est la même pour toutes les classes  $\rightarrow$  inutile de la calculer pour trouver la classe qui aura la plus forte probabilité

$P(C) \sim$  fréquence de la classe  $C$  dans le jeu d'apprentissage

**Problème** :  $P(X|C)$  est rarement accessible en pratique (car trop de combinaisons)

exemple avec ciel (5 modalités), température (4 modalités), humidité (10 modalités), vent (2 modalités)

$\rightarrow 5 \times 4 \times 10 \times 2 = 400$  combinaisons possibles

Assomption **naïve** : indépendance des attributs

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times \dots \times P(x_k | C)$$

si le  $i$ -ème attribut est catégoriel (variable qualitative) :

→  $P(x_i | C)$  est estimée comme la fréquence des échantillons qui ont pour valeur  $x_i$  et qui sont de classe  $C$

si le  $i$ -ème attribut est continu (variable quantitative) :

→  $P(x_i | C)$  est estimée avec une gaussienne (moyenne et écart-type des valeurs de l'attribut pour cette classe)

calcul facile

## Illustration

observations

Ciel	Température	Humidité	Vent	Classe
dégagé	chaud	forte	faux	<b>N</b>
dégagé	chaud	forte	vrai	<b>N</b>
couvert	chaud	forte	faux	<b>P</b>
pluvieux	moyenne	forte	vrai	<b>P</b>
pluvieux	frais	normale	faux	<b>P</b>
pluvieux	frais	normale	vrai	<b>N</b>
couvert	frais	normale	vrai	<b>P</b>
dégagé	moyenne	forte	faux	<b>N</b>
dégagé	frais	normale	faux	<b>P</b>
pluvieux	moyenne	normale	faux	<b>P</b>
dégagé	moyenne	normale	vrai	<b>P</b>
couvert	moyenne	forte	vrai	<b>P</b>
pluvieux	chaud	normale	faux	<b>P</b>
pluvieux	moyenne	forte	vrai	<b>N</b>

→ table de probabilités

Classe	
P( P ) = 9 / 14	P( N ) = 5 / 14
Ciel	
P(dégagé   P) = 2/9	P(dégagé   N) = 3/5
P(couvert   P) = 4/9	<b>P(couvert   N) = 0</b>
P(pluvieux   P) = 3/9	P(pluvieux   N) = 2/5
Température	
P(chaud   P) = 2/9	P(chaud   N) = 2/5
P(moyenne   P) = 4/9	P(moyenne   N) = 2/5
P(frais   P) = 3/9	P(frais   N) = 1/5
Humidité	
P(forte   P) = 3/9	P(forte   N) = 4/5
P(normale   P) = 6/9	P(normale   N) = 1/5
Vent	
P(vrai   P) = 3/9	P(vrai   N) = 3/5
P(faux   P) = 6/9	P(faux   N) = 2/5

objet X = <pluvieux, chaud, forte, faux>

$$P(X|P) \cdot P(P) = P(\text{pluvieux}|P) \cdot P(\text{chaud}|P) \cdot P(\text{forte}|P) \cdot P(\text{faux}|P) \cdot P(P)$$

$$= \frac{3}{9} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} = 0.010582$$

$$P(X|N) \cdot P(N) = P(\text{pluvieux}|N) \cdot P(\text{chaud}|N) \cdot P(\text{forte}|N) \cdot P(\text{faux}|N) \cdot P(N)$$

$$= \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} = \mathbf{0.018286}$$

→ X est classé comme N

**remarque** : P(X) n'a pas été utilisée, les valeurs obtenues ne sont donc pas des probabilités

... rend le calcul possible

... mène à des classificateurs optimaux si elle est vérifiée

... mais c'est rarement le cas, car les attributs sont souvent corrélés  
tentative de pallier cette difficulté :

- réseaux bayésiens : combinent le raisonnement bayésien avec la relation causale entre les attributs
- arbres de décision : considère un seul attribut à la fois, en commençant par le plus important

# Réseaux bayésiens

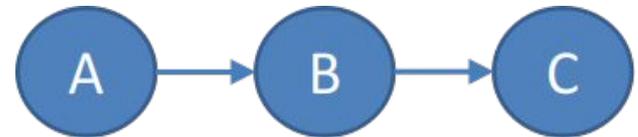
Motivation : les attributs ne sont pas indépendants

Solution : prise en compte partielle par un modèle réseau

exemple : modélisation de l'influence d'un facteur de transcription sur ces cibles

Le niveau d'expression de B dépend de celui de A (influence)

$P(B/A)$



Le niveau d'expression de C

- dépend de celui de B
- est indépendant de celui de A sachant B

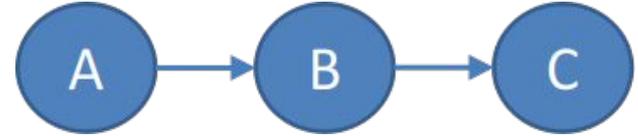
	cond1	cond2	cond3	cond4	cond5
A	off	low	high	low	low
B	off	low	high	high	low
C	off	off	low	high	low

observations

		A		
		off (1)	low (3)	high (1)
B	off	1/1	0/3	0/1
	low	0/1	2/3	0/1
	high	0/1	1/3	1/1

$P(B/A)$

Le niveau d'expression de B dépend de celui de A (influence)  
 $P(B/A)$



Le niveau d'expression de C

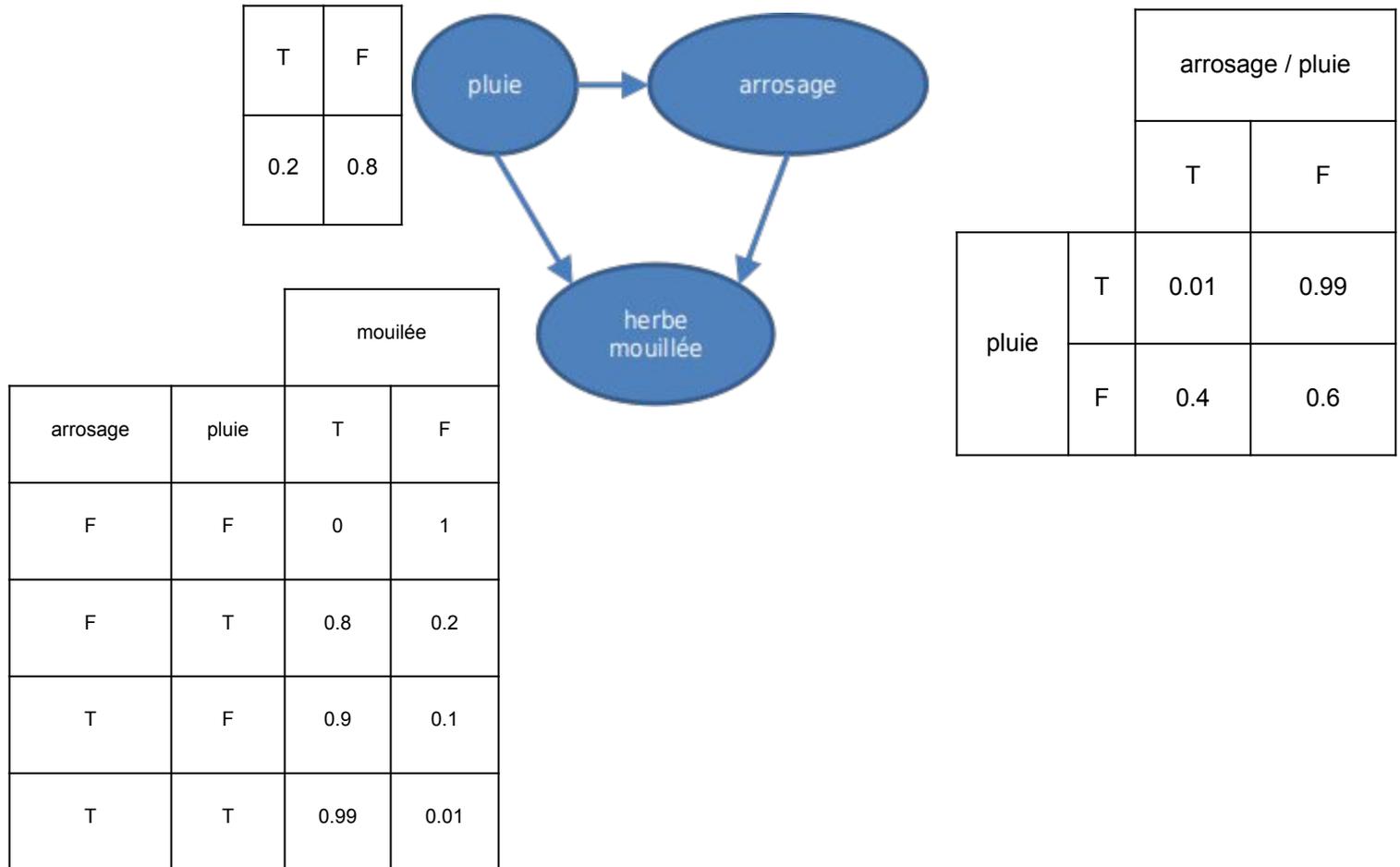
- dépend de celui de B
- est indépendant de celui de A sachant B

Avantages :

- capture l'aspect stochastique de la régulation
- possibilité d'intégrer des régulations connues
- peu de sur-apprentissage et robustesse
- quantitatif (niveau d'expression) ou qualitatif (on/off)

Inconvénient : Pas de cycle (circuit) donc pas de boucle d'auto-régulation

## Réseaux bayésiens : illustration

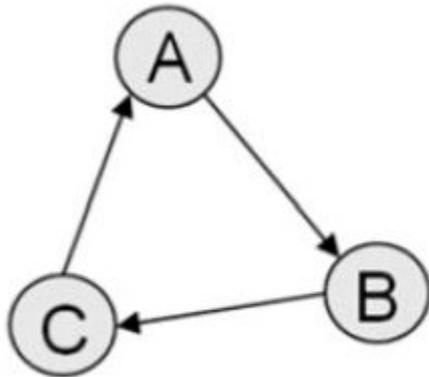


Permet de modéliser des boucles (d'autorégulation)

Information temporelle (temps discret)

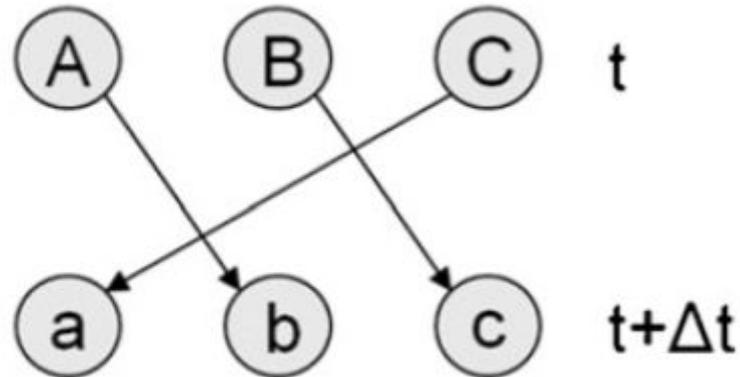
- variable indicée par son pas de temps
- distribution des probabilités d'une variable dépend de l'état de ses prédécesseurs au pas de temps précédent

Static BN



Not allowed !

Dynamic BN



source : Hecket *et al.* 2009