Décision dans l'Incertain

Olivier Pietquin
olivier.pietquin@univ-lille1.fr

Université Lille 1 - LIFL (SequeL)

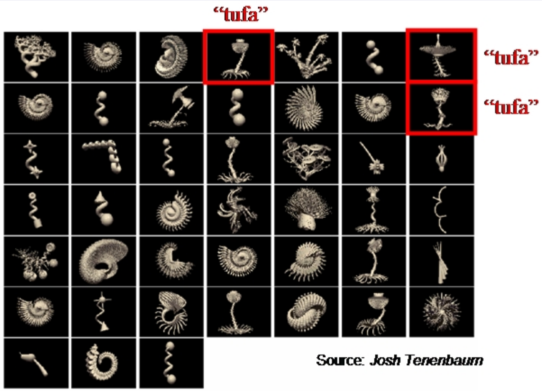
Master 2 Info - MOCAD

Quelques Problèmes I

Tout n'est pas dans les données

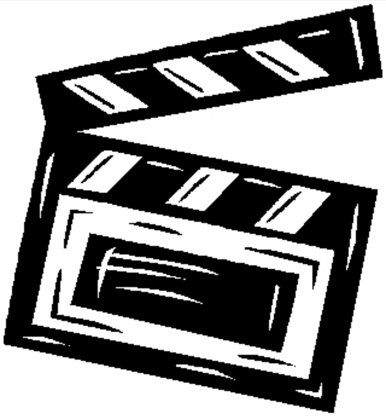
Quelques Problèmes II

Data sparsity



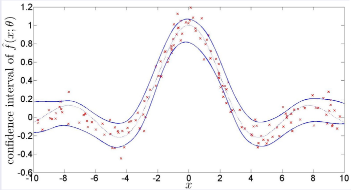
Quelques Problèmes III

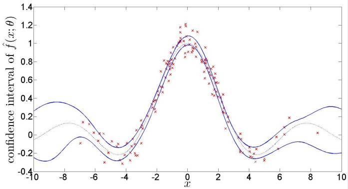
Apprentissage en ligne



Quelques Problèmes IV

Gérer l'incertitude (apprentissage actif)



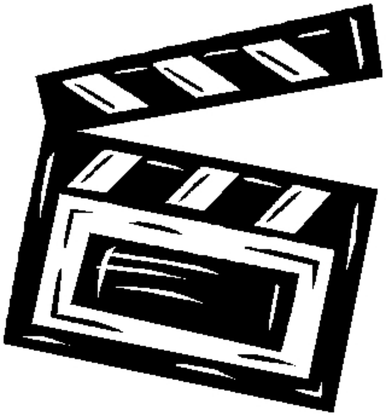


Quelques Problèmes V

Agir dans l'incertain

Quelques Problèmes VI

Contrôle optimal de systèmes complexes



Types de décision I

Décision de test

A partir d'observations générées par un environnement, l'agent prend une décision sur l'état de cet environnement.

Exemples

- Classification
- Détection
- Diagnostic
- Inférence dans les réseaux bayésiens.

Types de décision II

Décision d'action

A partir d'une succession d'actions et d'observations, l'agent prend une décision sur l'action à accomplir à l'instant présent. Cette action aura le plus souvent pour effet de modifier l'environnement et de générer une nouvelle observation.

Exemples

- Contrôle optimal
- Jeux

Types de décision III

En pratique ...

... ce n'est pas si simple ! Chaque décision à une composante de test (identification de l'état) et une composante d'action (en tirer les conséquences)

Exemples

- Diagnostic suivi d'un traitement
- Observation topologique en vision par ordinateur et action d'un robot

Position du problème général

Etats, actions, observations

Un **agent** est immergé dans un **environnement**. Il reçoit des **observations** E générées par cet environnement, qu'on appelle aussi **perceptions**. En utilisant ces perceptions, l'agent infère un **état** S probable de l'environnement (degré de croyance) et il doit réaliser une **action** A qui va l'amener vers un état $S_i(A, S)$ plus désirable dans l'environnement (désir, volonté, besoin).

Action

Une action est donc une fonction de l'état S $(A(S))$. A chaque action, plusieurs états de sortie $S_i(A, S)$ sont possibles avec chacun une probabilité $P(S_i|A, S)$.

Types d'apprentissage

Types d'apprentissage

Olivier Pietquin

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

Types d'apprentissage

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

Apprentissage Non-Supervisé

- Apprendre une structure dans un ensemble de données ;
- Pas d'oracle ;

Types d'apprentissage

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

Apprentissage Non-Supervisé

- Apprendre une structure dans un ensemble de données ;
- Pas d'oracle ;

Apprentissage par Renforcement

- Apprendre à se comporter !
- Apprentissage en ligne
- Décisions séquentielles

Structure du cours : 3 questions, 3 parties I

Incertitude et IA

Problème La logique formelle ne fonctionne qu'à partir de faits certains

Question Comment un agent intelligent peut-il tenir compte de son incertitude sur l'état du monde dans son processus de raisonnement ?

Réponse Probabilités

Exemples

Loi de Bayes, Règle de classification MAP, Algorithme d'apprentissage EM

Structure du cours : 3 questions, 3 parties II

Représentation et Manipulation

- **Problème** L'utilisation de la théorie des probabilités peut mener à une explosion combinatoire de la complexité algorithmique
- **Question** Comment représenter et manipuler efficacement les probabilités dans un agent intelligent ?
- **Réponse** Modèles graphiques capturant l'incertitude
- **But** Améliorer la connaissance statique du monde (apprentissage statique, classification)

Exemples

Réseaux Bayésiens (Dynamiques), Chaînes de Markov, Modèles de Markov Cachés, Filtres de Kalmann

Structure du cours : 3 questions, 3 parties III

Agir en Milieu Incertain

Problème Agir sur le monde peut influencer son état de manière incertaine, voire indésirable

Question Comment se comporter efficacement dans un milieu incertain ?

Réponse Contrôle optimal de systèmes stochastiques

But Créer des agents autonomes évoluant dans des environnements dynamiques

Exemples

Graphes de Décisions, Processus Décisionnels de Markov (Partiellement Observables), Programmation Dynamique, Apprentissage par Renforcement

Première partie I

IA Probabiliste ou Bayésienne

1 Introduction
- Définition et utilité
- Incertitude
- Approches des probabilités

2 Probabilités : Rappels

3 Classification Bayésienne

4 Apprentissage Bayésien

5 Apprentissage de paramètres

6 Régression Bayésienne

Intelligence Artificielle Probabiliste ?

Tentative de définition

Utilisation de la théorie des **probabilités** par un agent intelligent pour **apprendre** automatiquement, **raisonner** et **prendre des décisions** dans un domaine **incertain** à partir de **perceptions**.

Pourquoi ?

- Logique formelle : inférence logique sur base de faits certains
- Années 1960 : incapacité de la logique formelle à traiter le problème du diagnostic médical
- Nécessité de gérer l'incertain

Limites de la logique du 1er ordre

Beaucoup de problèmes pratiques ne trouvent pas de solution optimale par déductions logiques à partir des seules observations disponibles pour un agent.

Exemple 1 : diagnostic médical

- Symptômes de méningite : douleur dans la nuque
- Observation : douleur dans la nuque
- Conclusion : patient atteint de méningite ?

Exemple 2 : prédiction météo

- Paramètres : hauteur nuages, (gradient de) température, pression, saturation en eau etc.
- Même en connaissant tous les paramètres exactement, on ne peut rien en déduire de manière certaine.

Limites des autres méthodes d'apprentissage I

Qualifier l'incertitude

La plupart des méthodes de classification issues de l'apprentissage numérique apporte rarement de notion de confiance dans la décision.

Apprendre en ligne

Les méthodes classiques d'apprentissage numérique nécessitent des adaptations parfois complexes pour permettre un apprentissage prenant en compte des observations arrivant de manière continue.

Déformation locale

Les déformations locales (séquences temporelles ou spatiales) sont assez difficilement traitées par les autres méthodes.

Limites des autres méthodes d'apprentissage II

Connaissances *a priori*

Comment tenir compte de connaissance *a priori* sur le problème pour ne pas partir de rien et ne pas trop tenir compte des données quand on en a peu.

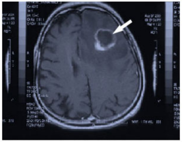
Causes de l'incertitude I

Conditions de reproductbilité

(Presque ?) Tous les événements sont **certains** mais les conditions de reproductibilité peuvent être complexes à maîtriser et impossibles à réaliser pratiquement (ex : pile ou face).

Ignorance Théorique

Certaines conditions nécessaires à la reproductibilité d'événements sont inconnues (ex : causes du cancer ?)



Magnetic resonance imaging showing brain tumor (arrow)

Causes de l'incertitude II

Fainéantise

Même si toutes les conditions sont connues et reproductibles, cela peut être fastidieux et on préfère s'en affranchir si possible.

Tout est relatif

Il y a des notions qui peuvent être vagues ou relatives (ex : un homme courageux peut l'être plus ou moins)

Probabilité, incertitude, degré de croyance

Approche Bayésienne

- La théorie des probabilités aide à intégrer nos incertitudes (ou celles d'un agent) sur l'état du monde dans un processus de décision.
- Les probabilités peuvent représenter un degré de croyance pour un agent intelligent.

Exemples

- On a une chance sur un million de se faire mordre par une chauve-souris enragée.
- Il y a 80% de chances pour qu'une douleur dentaire soit causée par une carie.

Approches des probabilités I

Approche Fréquentiste

La probabilité d'un événement est le **rapport** du nombre de cas favorables au nombre total d'événements.

$$P(x_i) = \lim_{n_{x_i} \to \infty} \frac{n_{x_i}}{\sum_j n_{x_j}}$$

Elle n'existe donc que par l'observation de l'événement et celui-ci doit donc être observable

Exemple

Lancer 100 fois ($\sum_j n_{x_j}$) un dé et compter le nombre de fois n_{x_i} où il s'arrête avec la face marquée d'un 1 vers le haut.

Approches des probabilités II

Approche Objectiviste

Les probabilités sont des **propriétés intrinsèques** de l'univers.

Exemple

- Certains phénomènes en physique quantique sont considérés comme intrinsèquement probabilistes.
- Dans le cas du dé, il s'agit plutôt des conditions initiales qui seraient aléatoires.

Remarque

Rem : L'approche fréquentiste (Venn) peut donc être considérée comme une méthode pour observer ces propriétés.

Approches des probabilités III

Approche Bayésienne ou Subjectiviste

Les probabilités représentent un **état de croyance** sur l'état du monde.

Exemple

- Nous sommes sur à 100% que les planètes d'un système solaire tournent autour de leur soleil.
- On estime à 1/1 000 000 les risques de perdre 2 moteurs lors d'une traversée de l'atlantique en avion.

Intérêts de l'approche Bayésienne

Compatibilité

- Si on dispose des résultats d'observations (**approche fréquentiste**) on peut s'en servir pour initialiser ou mettre à jour l'état de croyance.
- La théorie (**approche objectiviste**) peut aussi permettre de mettre à jour l'état de croyance.

Observations ?

- Les observations ne sont pas nécessaires au départ (pas besoin de générer une infinité de systèmes solaires pour avoir une certitude).
- Les observations permettent de mettre à jour l'état de croyance (ex : Twin Towers)

1 Introduction

2 Probabilités : Rappels
- Définitions et Axiomes
- Utilisation
- Inférence et probabilité jointe
- Loi de Bayes
- Historique

3 Classification Bayésienne

4 Apprentissage Bayésien

5 Apprentissage de paramètres

Définitions I

- Ω **: domaine** ou ensemble des événements possibles.
- $\mathbf{x} \in \Omega$: **événement** (élémentaire).
- $\mathbf{X}$: **variable aléatoire** qui prend ses valeurs dans Ω.
- $\mathbf{P(X = x)}$: **probabilité** que la variable aléatoire X prenne la valeur x c'est à dire que l'événement x se réalise. On note aussi $\mathbf{P(x)}$, **probabilité de l'événement** x.
- $P(X)$: probabilité **marginale** ou **a priori** de la variable X, désigne un ensemble d'équations $P(X = x)$ pour tous les $x \in \Omega$.

Définitions II

Pour 2 variables aléatoires A et B prenant leurs valeurs dans des ensembles Ω_A et Ω_B :

- $P(A \cup B)$ ou $P(A \vee B)$: probabilité de A **ou** B.
- $P(A \cap B), P(A, B)$ ou $P(A \wedge B)$: probabilité de A **et** B appelée **probabilité jointe** de A et B.
- $P(A|B) = \frac{P(A,B)}{P(B)}$: **probabilité conditionnelle** de A sachant B ou **a posteriori** (car disponible après avoir connaissance de B).
- A et B sont **indépendantes** si $P(A|B) = P(A)$ et donc $P(A, B) = P(A) \cdot P(B)$.

Remarque : Les notations impliquant des lettres majuscules désignent toujours un ensemble d'équations.

Définitions III

Un petit exercice :

	Total	Vaccinés (V)
Malades (M)	300	120
Sains (S)	700	480

Calculer : $P(M)$, $P(\bar{V})$, $P(M \cap V)$, $P(M|V)$, $P(M \cup V)$

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

Définitions IV

Indépendance conditionnelle :

- a et b sont conditionnellement indépendants sachant C si $P(a, b|C) = P(a|C)P(b|C)$.
- Attention : a et b sont des événements et C une variable. Définition plus forte : $P(A, B|C) = P(A|C)P(B|C)$.
- Exemple : angine, fièvre, mal de gorge.

Définitions V

3 types de variables aléatoires :

- **Variables booléennes** : domaine $= [vrai, faux]$
- **Variables discrètes** : domaine énumérable contenant des événements discrets mutuellement indépendants.
- **Variables continues** : domaine de nombres réels.

Représentation :

- **Variables discrètes** : vecteur de scalaires
- **Variables continues** : distributions de probabilités paramétrées
- **Probabilités conditionnelles** : Tables (CPT). Attention au mélange continu/discret.

Axiomes de Kolmogorov

Axiome

$$\forall x \in \Omega : P(x) \in [0,1]$$

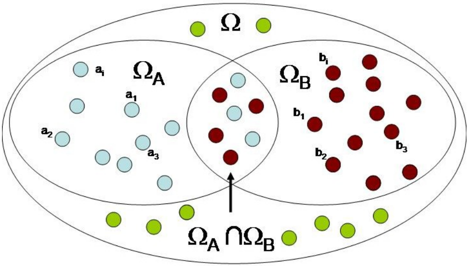
Axiome

$$P(\Omega) = 1 \ et \ P(\emptyset) = 0$$

Axiome

$$\forall \Omega_i, \Omega_j \in \Omega$$
$$Si \ \Omega_i \cap \Omega_j = \emptyset, \forall i \neq j : P(\cup_i \Omega_i) = \sum_i P(\Omega_i)$$
$$\sigma\text{-}additivit\acute{e}$$

Conséquences



Lemme

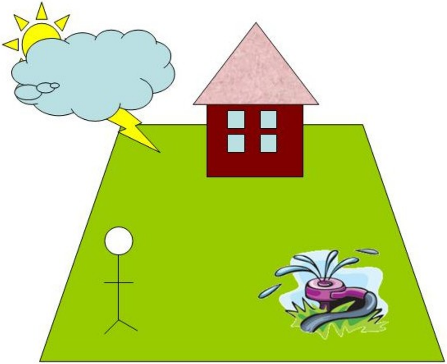
$$P(\Omega_A) = P(\cup_i a_i) = \sum_i P(a_i)$$

Lemme

$$P(\Omega_A \cup \Omega_B) = P(\Omega_A) + P(\Omega_B) - P(\Omega_A \cap \Omega_B)$$

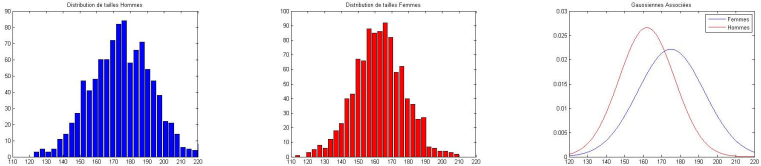
Utilisation en IA I

Inférence statistique causale.
Cas de l'arrosage automatique



Probabilité que le jardin soit humide étant donné que le ciel est nuageux : $P(humide|nuages)$

Classification
Genre $\Leftrightarrow$ connaissance de la taille en cm ?

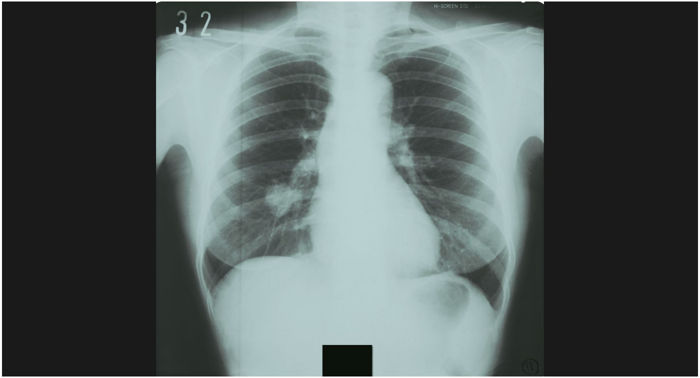


Probabilité d'observer un homme étant donné la mesure de la taille : $P(homme|Taille = 172cm)$

MOCAD - DI
Olivier Pietquin
Introduction
Probabilités : Rappels
Définitions et Axiomes
Utilisation
Inférence et probabilité jointe
Loi de Bayes
Historique
Classification Bayésienne
Apprentissage Bayésien
Apprentissage de paramètres
Régression Bayésienne

Utilisation en IA III

Diagnostic



Probabilité d'un cancer étant donné le cliché : $P(cancer|cliche)$

Utilisation en IA IV

Apprendre à agir

Probabilité d'avoir une meilleure donne étant donnée la donne actuelle ?
$P(prendre|donne)$

Probabilité d'obtenir un double 6 ?
$P(lancer|resultat)$

Inférence et Marginalisation

Marginalisation

La probabilité marginale ou *a priori* d'un événement est la somme des probabilités d'occurrence de cet événement conjointement avec tous les autres :

$$P(Y) = \sum_{z_i} P(Y, z_i) = \sum_{z_i} P(Y|z_i)P(z_i)$$

Inférence

L'inférence consiste à trouver la probabilité d'occurrence d'événements étant donnée l'observation de l'occurrence d'autres événements : $P(X|e)$

$$P(X|e) = \frac{P(X, e)}{P(e)} = \frac{\sum_{z_i} P(X, e, z_i)}{\sum_{z_i} P(e, z_i)}$$

Diagnostic

Problème du diagnostic

En général on cherche la probabilité d'une cause après observation d'un effet ou un symptôme. Or, on connaît plus souvent les effets ou les symptômes provoqués par les causes.

Exemple

On sait qu'une méningite provoque un mal de nuque. Si on observe un mal de nuque, doit on diagnostiquer une méningite ?

Loi de Bayes I

Par défintion

$$P(e,h) = P(e|h)P(h)$$
$$P(e,h) = P(h|e)P(e)$$

Loi de Bayes

$$\underbrace{P(h|e)}_{a\ \ posteriori} = \frac{\overbrace{P(e|h)}^{vraisemblance} \cdot \overbrace{P(h)}^{a\ \ priori}}{\underbrace{P(e)}_{a\ \ priori}}$$

Remarque

Aussi appelée Règle de Bayes ou Théorème de Bayes.

Loi de Bayes II

Autres formes

$$P(h|e) = \frac{P(e|h)P(h)}{\sum_i P(e|h_i)P(h_i)}$$

$$P(h|e,b) = \frac{P(e|h,b)P(h|b)}{P(e|b)}$$

Remarque

- *$P(e)$ peut être vu comme une normalisation.*
- *La première réécriture ramène à la vision fréquentiste (cas favorables sur somme des cas possibles)*
- *b : background*

Loi de Bayes III

$$\text{Loi de Bayes : } P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

Interprétations

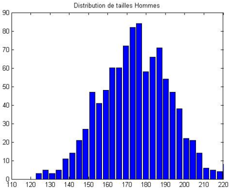
- **Diagnostic** : à partir de connaissances sur la causalité et sur la probabilité a priori d'événements, on peut inférer une cause à partir d'effets et donc raisonner dans l'incertitude.
- **Mise à jour** de l'état de croyance sur le monde : à partir de l'observation d'un événement, on peut changer l'état croyance en conséquence.

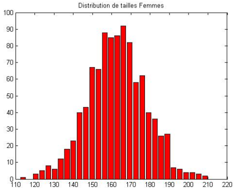
Intérêt

$P(e|h)$ ne dépend que de h et e, alors que $P(h|e)$ dépend de toutes les classes !

Utilisation en classification et diagnostic I

Classification en genre selon la taille :





$$P(homme) = 0.48$$
$$P(180|homme) = 0.085$$

$$P(femme) = 0.52$$
$$P(180|femme) = 0.038$$

$$P(homme|180) = \alpha 0.085 * 0.48 = \alpha 0.0408$$
$$P(femme|180) = \alpha 0.038 * 0.52 = \alpha 0.0198$$

Conclusion : 180cm implique plutôt le genre masculin.

Utilisation en classification et diagnostic II

Diagnostic de la méningite :

Données

- Probabilité d'observer une méningite si pas d'épidémie : $1/50000$
- Probabilité d'observer une douleur à la nuque : 0,05
- Probabilité qu'une méningite provoque une douleur à la nuque : 0,5

Solution

$$P(menigite|douleur) = \frac{P(douleur|meningite)P(meningite)}{P(douleur)} = \frac{0.5*1/50000}{0.05} = 0.0002$$

rem : on a multiplié par 10 notre croyance sur l'état "méningite"

Utilisation en classification et diagnostic III

Diagnostic de la méningite en cas d'épidémie :

Données

- Probabilité d'observer une méningite si épidémie : 1/500
- Probabilité d'observer une douleur à la nuque : 0,05
- Probabilité qu'une méningite provoque une douleur à la nuque : 0,5

Solution

$$P(menigite|douleur, epidemie) = \frac{P(douleur|meningite, epidemie)P(meningite|epidemie)}{P(douleur|epidemie)} = \frac{0.5 * 1/500}{0.05} = 0.02$$

Historique I

- 1654 : Antoine Gombauld, dit Chevalier de Méré (écrivain, joueur) perd au jeu des sommes importantes et ne comprend pas pourquoi.
- Il fait part de plusieurs problèmes à Pascal.
- Pascal et Fermat entretiennent une correspondance sur la résolution de ces problèmes.
- Il s'agit de probabilités directes !

Blaise Pascal

Pierre de Fermat

Historique II

- 1763 : Parution posthume d'un essais de Thomas Bayes (révérend presbytérien) dans lequel les probabilités sont vues comme des degrés de croyance.
- 1820 : Simon de Laplace partage cette vision des probabilités
- 1860 : John Venn publie "The Logic of Chance" où les probabilités sont vues comme des fréquences d'occurrences (vue fréquentiste).

Historique III

Thomas Bayes (1702-1761)

Utilité et risque I

Prendre une décision = faire un acte utile

Lorsqu'un agent intelligent prend une décision c'est pour atteindre un but. La décision la plus appropriée est celle qui est la plus utile pour atteindre le but.

Utilité

On associe une **utilité** $u(h|f)$ à toute décision associant l'hypothèse h à l'événement e alors que l'hypothèse correcte était f.

L'utilité moyenne ou bayésienne associée à la classification de l'événement e est alors donnée par :

$$U(h|e) = \sum_f u(h|f)P(f|e)$$

Utilité et risque II

Prendre une décision = prendre un risque

Toute prise de décision dont les conséquences sont inconnues constituent une prise de risque. La décision la plus appropriée est souvent considérée comme la moins risquée.

Risque

On associe un **coût** $c(h|f)$ à toute décision associant l'hypothèse h à l'événement e alors que l'hypothèse correcte était f.
Le risque moyen ou bayésien associé à la classification de l'événement e est alors donné par :

$$R(h|e) = \sum_{f} c(h|f)P(f|e)$$

Utilité et risque III

Décision

La décision optimale est donc donnée par :

$$h^* = \underset{h}{\operatorname{argmax}}\, U(h|e) \;\; si \max_h U(h|e) > \sigma_U$$

$$h^* = \underset{h}{\operatorname{argmin}}\, R(h|e) \;\; si \min_h R(h|e) < \sigma_R$$

Sinon : Rejet !
rem : on définit un seuil d'utilité σ_U ou de risque σ_R auquel est associé une décision de rejet.

Utilité et risque IV

Utilisation de la règle de Bayes

$$h^* = \underset{h}{\operatorname{argmax}}\, U(h|e) = \underset{h}{\operatorname{argmax}} \sum_f u(h|f) \frac{P(e|f)P(f)}{P(e)}$$

$$h^* = \underset{h}{\operatorname{argmax}}\, U(h|e) = \underset{h}{\operatorname{argmax}} \sum_f u(h|f) P(e|f)P(f)$$

Théorie de la décision

L'introduction de la notion d'utilité (ou de risque) donne lieu à la **théorie de l'utilité**. L'association de cette théorie et de celle des probabilités donne lieu à la **théorie de la décision**.

$$\text{Decision theory} = \text{Probability theory} + \text{Utility theory}$$

Règle du Maximum A Posteriori : MAP

Utilité identique et nulle pour toutes les hypothèses erronées.
Soit $u(h|f) = 0$ si $h \neq f$ et $u(h|f) = 1$ sinon.

$$h^* = \underset{h}{\operatorname{argmax}}\, U(h|e) = \underset{h}{\operatorname{argmax}} \sum_{h} u(h|f)P(e|f)P(f)$$

Règle du Maximum a Posteriori - MAP

$$h^* = \underset{h}{\operatorname{argmax}}\, P(e|h)P(h) = \underset{h}{\operatorname{argmax}}\, P(h|e)$$

On choisit donc l'hypothèse qui a la probabilité *a posteriori* maximale

Remarque

- *Cela fonctionne de la même manière avec le risque*
- *L'hypothèse reste forte (même coût pour toutes les erreurs, faux en diagnostic médical) mais est souvent utilisée.*

Maximum de vraisemblance ou ML

Si les hypothèses ont toutes la même probabilité *a priori* : $P(h_i) = P(h) \forall h_i$.

Règle du Maximum de Vraisemblance - ML

$$h^* = \underset{h}{\operatorname{argmax}} P(e|h)$$

- On choisit l'hypothèse pour laquelle la vraisemblance de l'événement est maximale.
- in English : Maximum Likelihood (**ML**)
- Rem : on remplace tout simplement $P(h|e)$ par $P(e|h)$!
- ex : reconnaissance de chiffres écrits

Classificateur naïf de Bayes I

Souvent, on dispose de plusieurs observations en même temps. On cherche alors $P(h|e_1, e_2, \ldots, e_n)$

Règle de Bayes : $P(h|e_1, e_2, \ldots, e_n) = \frac{P(e_1, e_2, \ldots, e_n|h)P(h)}{P(e_1, e_2, \ldots, e_n)}$

Si les événements e_i sont conditionnellement indépendants sachant l'hypothèse h on : $P(e_1, e_2, \ldots, e_n|h) = \prod_i P(e_i|h)$

Classificateur naïf de Bayes

$$h^* = \operatorname*{argmax}_h P(h) \prod_i P(e_i|h)$$

Classificateur naïf de Bayes II

ex : diagnostic de la grippe

Symptômes : fièvre, mal de tête, douleur à la gorge.

$$P(grippe|fievre, tete, gorge) = \frac{P(fievre, tete, gorge|grippe)P(grippe)}{P(fievre, tete, gorge)}$$

$$= \frac{P(fievre|grippe)P(tete|grippe)P(gorge|grippe)P(grippe)}{P(fievre, tete, gorge)}$$

Ces valeurs sont beaucoup plus simples à mesurer, mais il faut faire l'hypothèse que la fièvre n'est pas la cause du mal de tête. C'est la raison pour laquelle il est appelé *classificateur naïf*.

Apprentissage Bayésien I

Tentative de définition

L'apprentissage Bayésien consiste à faire usage de la règle de Bayes pour améliorer l'état de croyance d'un agent pour qu'il reflète mieux la réalité à partir d'observations ou de données.

Formellement

- Données ou exemples $\mathbf{D} = \{D_i = d_i\}$ avec $\mathbf{d} = \{d_i\}$
- Ensemble d'hypothèses $\mathcal{H} = \{h_i\}$ sur le processus qui a généré $\mathbf{D}$
- Loi de Bayes :

$$P(h_i|\mathbf{d}) = \frac{P(\mathbf{d}|h_i)P(h_i)}{P(\mathbf{d})} = \frac{P(\mathbf{d}|h_i)P(h_i)}{\sum_{h_k \in \mathcal{H}} P(\mathbf{d}|h_k)P(h_k)}$$

Apprentissage Bayésien II

Clés de l'apprentissage

- Probabilités *a priori* des hypothèses $P(h_i)\forall i$
- Vraisemblance des données sachant les hypothèses $P(\mathbf{d}|h_i)\forall i$

Hypothèse i.i.d.

On fait souvent l'hypothèse que les exemples sont indépendants et identiquement distribués :

$$P(\mathbf{d}|h_i) = \prod_j P(d_j|h_i)$$

$$P(h_i|\mathbf{d}) = \frac{P(\mathbf{d}|h_i)P(h_i)}{P(\mathbf{d})} = \frac{\prod_j P(d_j|h_i)P(h_i)}{\sum_{h_k \in \mathcal{H}} \prod_j P(d_j|h_k)P(h_k)}$$

Prédiction I

L'agent peut se servir de son nouvel état de croyance pour inférer la valeur la plus probable d'une variable inconnue X.

Prédiction Bayésienne

$$P(X|\mathbf{d}) = \sum_i P(X, h_i|\mathbf{d}) = \sum_i P(X|\mathbf{d}, h_i)P(h_i|\mathbf{d})$$

$$P(X|\mathbf{d}) = \sum_i P(X|h_i)P(h_i|\mathbf{d})$$

rem : les hypothèses apparaissent comme des intermédiaires de calcul !

Règle MAP

$$P(X|\mathbf{d}) = P(X|h_{MAP})$$

Exemple : paquet de bonbons

Dans un paquet de bonbons, il y en a de 2 parfums (cerise et citron vert) mais ils sont emballés de la même manière. Il existe 5 types de paquet contenant des proportions différentes. Je prends 10 bonbons les uns après les autres dans le paquet. Si je ne connais pas le type du paquet comment prévoir la probabilité que le prochain bonbon tiré du paquet soit au citron vert ?

On suppose que le paquet est assez grand pour que les tirages soient **i.i.d.**

Prédiction III

Données

Types de paquet et probabilités *a priori* :

- h_1 : 100% de cerises, $P(h_1) = 0.1$
- h_2 : 75% cerises et 25% citron, $P(h_2) = 0.2$
- h_3 : 50% cerises et 50% citron, $P(h_3) = 0.4$
- h_4 : 25% cerises et 75% citron, $P(h_4) = 0.2$
- h_5 : 100% citron, $P(h_5) = 0.1$

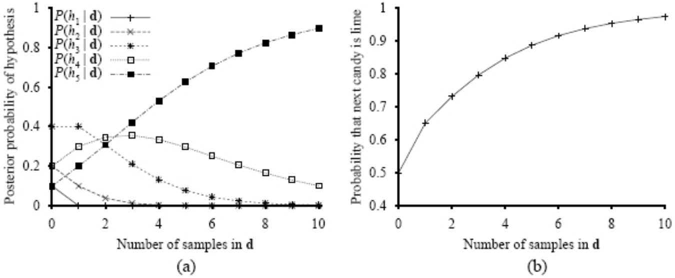
On tire 10 bonbons d'affiler, tous au citron.

On calcule :

$$P(h_i|citron) = \frac{P(citron|h_i) * P(h_i)}{\sum_i P(citron|h_j)P(h_j)}$$

$$P(citron|citron) = \sum_j P(X|h_j)P(h_j|citron)$$

Prédiction IV



- Au début on commence au niveau des priors
- L'hypothèse h_5 voit sa probabilité augmenter au fur et à mesure
- Les autres hypothèses diminuent rapidement
- La prédiction d'un nouveau bonbon au citron augmente
- Après 3 observations, la règle MAP prédirait un nouveau citron avec une probabilité de 1 alors que la loi de Bayes donne une probabilité de 0.8 !

Apprentissage de paramètres

Paramètres ?

L'apprentissage Bayésien consiste à calculer des probabilités *a posteriori* (après observation d'événements) à partir de vraisemblance des événements connaissant le phénomène qui est susceptible de les avoir générés et de probabilité *a priori* sur ces événements.

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)} = \alpha P(e|h)P(h)$$

Les paramètres sont donc :

- La vraisemblance des événements : $P(e|h)$
- La probabilité *a priori* du phénomène : $P(h)$
- Eventuellement la probabilité *a priori* de l'événements : $P(e)$ ou α

Probabilité a priori $P(h)$

En l'absence de connaissance, on peut supposer toutes les hypothèses équiprobables :

$$P(h_i) = \frac{1}{n_h}$$

Si on dispose de données étiquetées on peut assigner la fréquence d'apparition de chaque hypothèse :

$$P(h_i) = \frac{n_{h_i}}{n_h}$$

Pour tenir compte de connaissance *a priori* et de données, on peut utiliser un comptage virtuel :

$$P(h_i) = \frac{n_{h_i} + \frac{M_i}{C}}{n_h + M_i}$$

Données complètes : Maximum de Vraisemblance I

Maximum de vraisemblance

On cherche les paramètres θ du modèle décrivant les hypothèses qui maximisent la vraisemblance des données d :

$$\mathcal{L}(d, \theta) = P(d|\theta) = \prod_i P(d_i|\theta)$$

En général on utilise plutôt le logarithme de la vraisemblance (*log-likelihood*) :

$$\log \mathcal{L}(d, \theta) = L(d, \theta) = \sum_i \log P(d_i|\theta)$$

Donc, on dérive et on annule :

$$\frac{\delta L(d, \theta)}{\delta \theta_i} = 0$$

Données complètes : Maximum de Vraisemblance II

Cas discret

On observe n tirages au jeu du *pile ou face*. On observe p occurrences de *pile* et f occurrences de *face*. Quelle est la fraction θ d'événements "*pile*" que peut générer la pièce ?

$$P(d|\theta) = \prod_i P(d_i|\theta) = \theta^p(1-\theta)^f$$

$$L(d|\theta) = \log \prod_i P(d_i|\theta) = \theta^p(1-\theta)^f = p\log\theta + f\log(1-\theta)$$

$$\frac{\delta L(d|\theta)}{\delta\theta} = \frac{p}{\theta} - \frac{f}{1-\theta} = 0 \Rightarrow \theta = \frac{p}{p+f} = \frac{p}{n}$$

Données complètes : Maximum de Vraisemblance III

Cas continu

On observe n femmes et on mesure leur taille t en cm. On associe une distribution gaussienne à la probabilité de cette mesure et on cherche les paramètres μ et σ de la gaussienne.

$$P(t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

$$L(t,\mu,\sigma) = \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t_i-\mu)^2}{2\sigma^2}} = N(-\log\sqrt{2\pi} - \log\sigma) - \sum_{i=1}^{N} \frac{(t-\mu)^2}{2\sigma^2}$$

$$\frac{\delta L(t,\mu,\sigma)}{\delta\mu} = -\frac{1}{2\sigma^2} \sum_{i=1} N(t_i-\mu) = 0 \Rightarrow \mu = \frac{\sum_{i=1}^{N} t_i}{N}$$

$$\frac{\delta L(t,\mu,\sigma)}{\delta\sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{N} (t_i-\mu)^2 \Rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{N}(t_i-\mu)^2}{N}}$$

Données Complètes : apprentissage Bayésien I

Problème

Après l'observation d'un seul lancé de la pièce, si elle tombe sur pile, doit-on pour autant penser que la fraction de "pile" est de 100% ? Plus formellement, la méthode ML fonctionne si le nombre d'exemples tend vers ∞. Comment tenir compte de notre connaissance a priori pour améliorer la prédiction ?

Données Complètes : apprentissage Bayésien II

Règle de Bayes

Si on lance une fois une pièce et qu'elle tombe sur pile :

$$P(\theta|p) = \alpha P(p|\theta)P(\theta) = \alpha\theta P(\theta)$$

Si on lance une 2ème fois et qu'elle tombe encore sur pile :

$$P(\theta|p,p) = \alpha P(p,p|\theta)P(\theta) = \alpha\theta^2 P(\theta)$$

Si on lance encore une fois et qu'elle tombe sur face :

$$P(\theta|p,p,f) = \alpha P(p,p,f|\theta)P(\theta) = \alpha\theta^2(1-\theta)P(\theta)$$

D'une manière générale :

$$P(\theta|e) = \alpha P(e|\theta)P(\theta) = \alpha\theta^{n_p}(1-\theta)^{n_f} P(\theta)$$

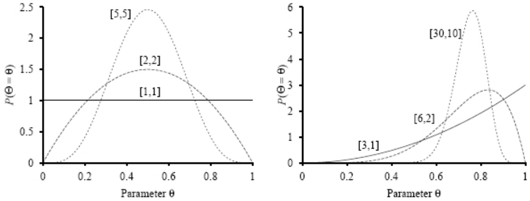
Données Complètes : apprentissage Bayésien III

Probabilité a priori sur θ

L'exemple précédent nécessite $P(\theta)$. On peut choisir n'importe quelle forme de distribution !

Distribution beta : $B(a, b)$

$$B(a, b) = P(\theta | a, b) = \alpha \theta^{a-1} (1 - \theta)^{b-1}$$



Données Complètes : apprentissage Bayésien IV

Conséquences

$$P(\theta|a,b,e) = \alpha\theta^{a-1+n_p}(1-\theta)^{b-1+n_f}$$

Les paramètres a et b apparaissent comme des tirages virtuels (a pour pile, b pour face). La moyenne de la distribution $P(\theta|a,b)$ est $\frac{a}{a+b}$ et plus $a+b$ est grand, plus la distribution est étroite autour de la moyenne.

On affecte donc les valeurs de a et b pour que la moyenne $\frac{a}{a+b}$ soit égale à la valeur présumée de θ et on les choisit aussi en fonction de notre certitude concernant cette moyenne.

$$(a,b) = (2,3) \Rightarrow E(\theta) = 0.4$$

$$(a,b) = (20,30) \Rightarrow E(\theta) = 0.4$$

Données Complètes : apprentissage Bayésien V

Remarque

- *La distribution de probabilité a posteriori d'une distribution beta est encore une distribution beta. On dit que c'est une famille a posteriori conjuguée (conjugate prior distributions).*
- *Pour des problèmes multivariés, on utilise la distribution de Dirichlet qui a la même propriété.*

Données incomplètes

Problème

Il arrive souvent que les données soient incomplètes, simplement parce qu'elles ne sont pas observables. On parle de données incomplètes, de **données manquantes** *ou de* **variables cachées***.*

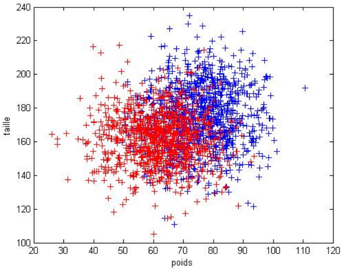
Exemples

- Liste de symptômes, mais pas les maladies correspondantes.
- Mesure et poids d'individus sans différencier le sexe.
- Etiquetage automatique d'une base de données en reconnaissance de forme (vocale).

Expectation-Maximization I

Cas simple : distribution multigaussienne

On suppose qu'on observe les effets de différentes causes et que chaque cause induit une distribution gaussienne de ses effets. Par exemple, on observe le poids et la taille d'individus sans connaître le sexe.



Expectation-Maximization II

Formalisation du problème

Dans le cas multigaussien (ou mélange de gaussiennes) on a :

$$P(\mathbf{x}) = \sum_i P(\mathbf{x}|C_i)P(C_i)$$

Il faut donc trouver les paramètres suivants :

- $\omega_i = P(C_i)$ (variables cachées)
- μ_i
- Σ_i

Expectation-Maximization III

Idée de solution

Si on connaissait les variables cachées, il suffirait de calculer le maximum de vraisemblance classique pour trouver les autres paramètres.
En initialisant ces paramètres à des valeurs quelconques, on peut obtenir une estimation des variables cachées et itérer !

Expectation-Maximization IV

Algorithme E-M

Initialisation

- Choisir une valeur initiale $\widehat{\theta} = \theta_0$ pour les paramètres.

Etape de calcul de l'espérance des variables cachées en utilisant l'estimation courante des paramètres $\widehat{\theta}$ (Etape E)

- Calculer $P(C_i|x, \widehat{\theta})$ avec la loi de Bayes

Etape de Maximisation (Etape M)

- Calculer les nouveaux paramètres qui maximisent la vraisemblance des données

$$\begin{gathered}\widehat{\mu_i} \leftarrow \sum_j p_{ij} x_j / p_i \\ \widehat{\Sigma_i} \leftarrow \sum_j p_{ij} x_j x_j^T / p_i \\ \widehat{\omega_i} \leftarrow p_i\end{gathered}$$

Expectation-Maximization V

EM formellement

Etape E : calcul de $P(\omega_i|x, \widehat{\theta_t})$

Etape M : Calcul de

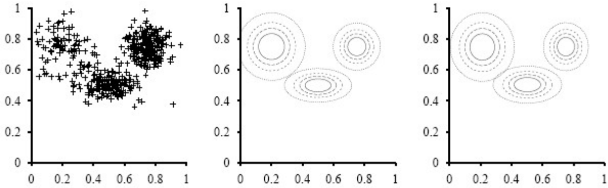
$$\widehat{\theta_{t+1}} = \underset{\theta}{\text{argmax}} \sum_i \log(P(x, \omega_i|\theta))P(\omega_i|x, \widehat{\theta_t})$$

Interprétation

On cherche à maximiser l'espérance de la vraisemblance des données complétées (Expectation Maximisation ou maximisation de l'espérance).

Expectation-Maximization VI

Exemple :



- Remarque : la vraisemblance est meilleure avec les valeurs apprises qu'avec les gaussiennes d'origine
- Attention à l'overfitting !

Analogue à une descente de gradient. Converge vers un minimum local

1 Introduction

2 Probabilités : Rappels

3 Classification Bayésienne

4 Apprentissage Bayésien

5 Apprentissage de paramètres

6 Régression Bayésienne
- Principe Général
- Régression linéaire
- Extensions
- Processus Gaussiens

Régression

Le problème de régression

Etant donné un ensemble $\mathcal{D}$ de données sous forme de couples $\{\mathbf{x}_i, y_i = f(x_i)\}$ (avec souvent $\mathbf{x}_i \in \mathbb{R}^m$ et $y_i \in \mathbb{R}$), il s'agit de trouver une fonction $\hat{f}(x)$ qui permette d'expliquer et de prédire la relation entre des entrées quelconques $(\mathbf{x})$ et les sorties (y) correspondantes.

Régression bayésienne paramétrique

Etant donné une famille de fonctions paramétriques f_θ, solutions potentielles du problème, trouver la distribution *a posteriori* $p(f_\theta|\mathcal{D})$ ou plus simplement $p(\theta|\{\mathbf{x}_i, y_i\})$.

Modèle linéaire simple

Modèle

La relation entre les $\mathbf{x}$ et les y est linéaire :

$$f_\theta(\mathbf{x}_i) = \sum_{j=1}^{m} \theta^j x_i^j = \theta^T \mathbf{x}_i$$

où c^j est la j$^{\text{ème}}$ composante du vecteur $\mathbf{c}$ et $\theta = \{\theta^1, \ldots, \theta^m\}^T$

Bruit gaussien

On supposera que l'erreur de modélisation ainsi que le bruit sur les données sont gaussiens :

$$y_i = f_\theta(\mathbf{x}_i) + \epsilon_i \text{ où } \epsilon_i \sim \mathcal{N}(0, \sigma_b^2)$$

Solution Bayésienne I

Reformulation

$$p(\theta|X, \mathbf{y}) = \frac{p(\mathbf{y}|\theta, X)p(\theta|X)}{p(\mathbf{y}|X)}$$

où y est le vecteur colonne des y_i et X est une matrice dont chaque colonne est un $\mathbf{x}_i$.
Puisque l'*a priori* est indépendant des données, le dénominateur est indépendant de θ et la solution est linéaire. On a :

$$p(\theta|X, \mathbf{y}) \propto p(\mathbf{y}|\theta^T X + \epsilon)p(\theta)$$

Il faut donc trouver la vraisemblance et l'*a priori*

Solution Bayésienne II

Vraisemblance

Puisque $\epsilon_i = y_i - \theta^T \mathbf{x}_i = \mathcal{N}(0, \sigma_b^2)$ et $\{\mathbf{x}_i, y_i\}$ iid on a :

$$
\begin{aligned}
p(\mathbf{y}|X, \theta) &= \prod_i \frac{1}{\sqrt{2\pi\sigma_b^2}} e^{-\frac{(y_i - \theta^T \mathbf{x}_i)^2}{2\sigma_b^2}} \\
&= \frac{1}{(2\pi\sigma_b^2)^{n/2}} e^{-\frac{1}{2\sigma_b^2}|\mathbf{y} - \theta^T X|^2} \\
&= \mathcal{N}(\theta^T X, \sigma_b^2 I)
\end{aligned}
$$

Solution Bayésienne III

Remarque : maximum likelihood

En résolvant le problème de maximum de vraisemblance on résout $\theta^* = \text{argmax}_\theta \log p(\mathbf{y}|X, \theta)$, soit :

$$\begin{aligned}
\theta^* &= \underset{\theta}{\text{argmin}} \sum_i (y_i - \theta^T \mathbf{x}_i)^2 \\
&= \underset{\theta}{\text{argmin}} (\mathbf{y} - \theta^T X)^T (\mathbf{y} - \theta^T X) \\
&= (XX^T)^{-1} X \mathbf{y}
\end{aligned}$$

Soit la solution des moindres carrés !

NB : les moindres carrés donnent donc la solution du maximum de vraisemblance pour un bruit gaussien centré.

Solution Bayésienne IV

A priori

On choisi un *a priori* gaussien sur les paramètres : $\theta \sim \mathcal{N}(\mathbf{0}, \Sigma_m)$

Solution

$$p(\theta|X,\mathbf{y}) \propto e^{-\frac{1}{2\sigma_b^2}(\mathbf{y}-\theta^T X)^T(\mathbf{y}-\theta^T X)} e^{\frac{1}{2}\theta^T \Sigma_m \theta}$$
$$\propto e^{-\frac{1}{2}(\theta-\bar{\theta})^T(\frac{1}{\sigma_b^2}X^T X+\Sigma_m^{-1})(\theta-\bar{\theta})}$$

où $\bar{\theta} = \frac{1}{\sigma_b^2}(\frac{1}{\sigma_b^2}XX^T + \Sigma_m^{-1})^{-1}X^T\mathbf{y}$
ou encore $\bar{\theta} = \frac{1}{\sigma_b^2}A^{-1}X^T\mathbf{y}$ où $A = \frac{1}{\sigma_b^2}XX^T + \Sigma_m^{-1}$. Et donc :

$$p(\theta|X,\mathbf{y}) \propto \mathcal{N}(\bar{\theta}, A^{-1})$$

Liens avec d'autres solutions

ML et MAP

On a vu que la solution des moindres carrés est la même que celle du maximum de vraisemblance :

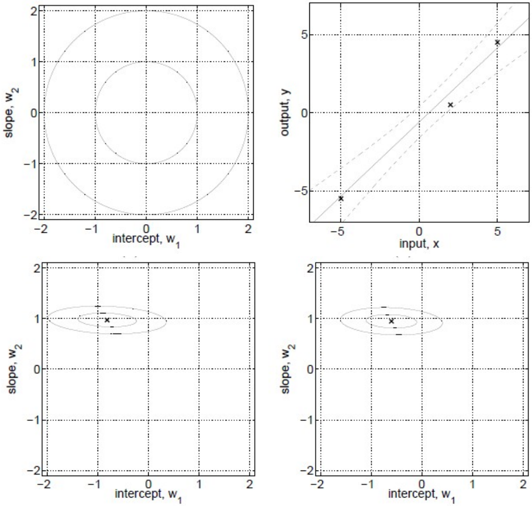
$$\theta^*_{ML} = (XX^T)^{-1}X\mathbf{y}$$

La solution du maximum *a posteriori* pour une solution gaussienne est donnée par le vecteur moyenne, ici c'est $\bar{\theta}$:

$$\theta^*_{MAP} = \frac{1}{\sigma_b^2}(\frac{1}{\sigma_b^2}XX^T + \Sigma_m^{-1})^{-1}X^T\mathbf{y}$$

NB : tout se passe comme si le *prior* introduisait un terme de pénalisation $\frac{1}{2}\theta^T\Sigma_m^{-1}\theta$

Exemple : $y = f(x) = w_1 + w_2 x$



Extensions

Régression non-linéaire

Pour régresser des fonctions non-linéaires :

$$f(x) = \theta^T \phi(x)$$

Avec $\phi(x)$ une fonction non-linéaire de x. Développements nécessitent le calcul de produits scalaires $\Rightarrow$ Kernel trick.

Processus Gaussiens

Définition

Extension des vecteurs gaussiens aux fonctions :

$$f(\mathbf{x}) \propto \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \text{ avec}$$
$$m(\mathbf{x}) = E[f(\mathbf{x})],$$
$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

Une fonction peut être vue comme un vecteur infini. Chaque composante $f(\mathbf{x})$ suit une loi gaussienne. La moyenne est calculée pour un $\mathbf{x}$ donné !

Régression non-linéaire

Si $f(\mathbf{x}) = \theta^T \phi(\mathbf{x})$ avec $\theta \sim \mathcal{N}(\mathbf{0}, \Sigma_m)$ on a :

$$E[f(\mathbf{x})] = E[\theta]^T \phi(\mathbf{x}) = \mathbf{0}$$
$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T E[\theta\theta^T]\phi(\mathbf{x}) = \phi(\mathbf{x})^t \Sigma_m \phi(\mathbf{x})$$

Régression avec GP

Sans bruit

Si on note comme précédemment $\{X, \mathbf{y} = \mathbf{f}\}$ les données d'entraînement et $\{X_*, \mathbf{y}_* = \mathbf{f}_*\}$ les points sur lesquels on veut connaître la fonction régressée on a :

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

Avant entraînement (*prior*) :

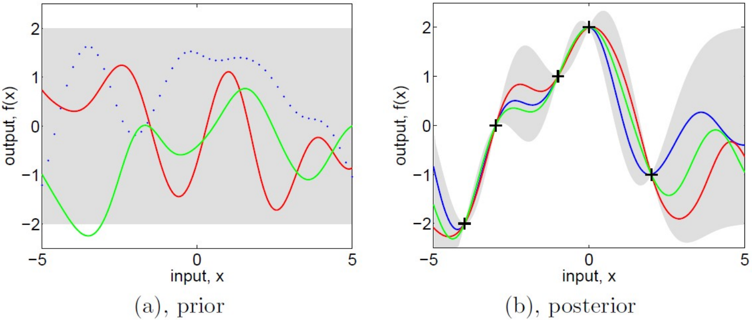
$$f_* = \mathcal{N}(\mathbf{0}, K(X_*, X_*))$$

Après entraînement (*posterior*) :

$$\begin{aligned} \mathbf{f}_* | X, X_*, \mathbf{f} \sim & \mathcal{N}(K(X_*, X)K(X,X)^{-1}\mathbf{f}, \\ & K(X_*, X_*) - K(X_*, X)K(X,X)^{-1}K(X, X_*)) \end{aligned}$$

Exemple

$$k(\mathbf{x}, \mathbf{x}') = e^{-\frac{1}{2}|\mathbf{x}-\mathbf{x}'|^2}$$



Deuxième partie II

Modèles Graphiques pour l'Inférence

Besoin de modèles

Pourquoi ?

Pour utiliser les méthodes décrites il faut pouvoir calculer des probabilités marginales et conditionnelles. On a vu que la probabilité jointe permettait de tout calculer :

$$P(e_i|h_j) = \frac{P(e_i, h_j)}{P(h_j)} = \frac{\sum_{\mathbf{z}} P(e_i, h_j, \mathbf{z})}{\sum_{i,\mathbf{z}} P(e_i, h_j, \mathbf{z})}$$

mais c'est un peu calculatoire et cela nécessite beaucoup de mémoire (CPT complète pour n variables booléennes : 2^n).

On a donc besoin de modèles pour exprimer le plus simplement possible la probabilité jointe.

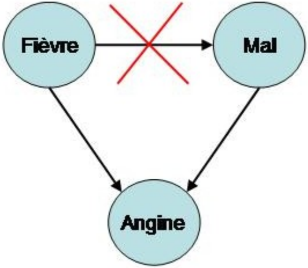
Indépendance conditionnelle

Rappel

- a et b sont conditionnellement indépendants sachant C si $P(a, b|C) = P(a|C)P(b|C)$.
- Attention : a et b sont des événements et C une variable. Définition plus forte : $P(A, B|C) = P(A|C)P(B|C)$.

Intérêt

L'indépendance conditionnelle permet de factoriser la probabilité jointe et d'estimer des paramètres plus simples : P(fievre,mal de gorge—angine) = P(fievre—C) P(mal de gorge—angine)



Réseaux Bayésiens : définition

Définition

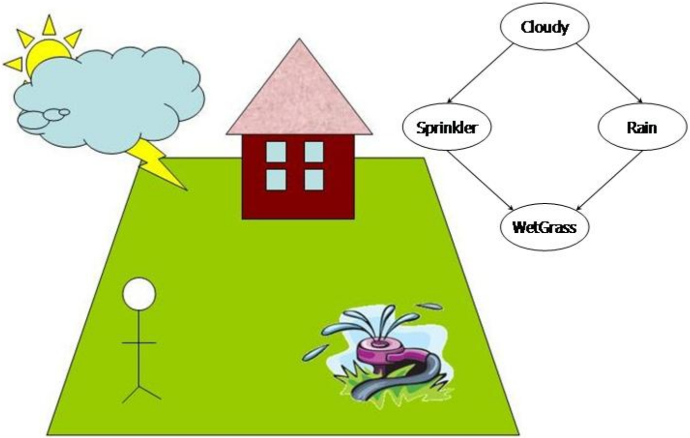
Réseau Bayésien

- *Un réseau bayésien est un graphe orienté acyclique (DAG : directed acyclic graph) dont les noeuds sont des variables aléatoires*
- *Si un arc existe entre X et Y alors, X est un parent de Y.*
- *Chaque noeud X_i possède une distribution conditionnelle $P(X_i|Parents(X_i))$*
- *Un arc est tiré de X vers Y si X influence directement Y.*

Vocabulaire

- Réseau (d'inférence) bayésien ou Réseau de croyance
- Bayesian Network ou Belief Network

Exemple : arrosage automatique (Murphy)



Factorisation de la loi jointe

Probabilité conditionnelle

$$P(x_1, x_2, ..., x_n) = P(x_n|x_1, x_2, ..., x_{n-1})P(x_1, x_2, ..., x_{n-1})$$

Chain Rule : $P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_{i-1}, ..., x_1)$
Dans un réseau bayésien :

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

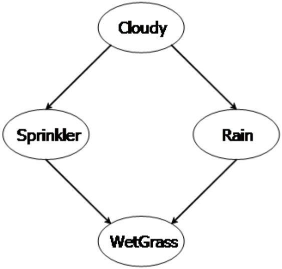
Si $Parent(X_i)) \subseteq (X_{i-1}, ..., X_1)$ le réseau décrit parfaitement une factorisation de la loi jointe

Intérêt

- Encoder les indépendances conditionnelles entre variables aléatoires
- Factorisation simplifiée d'une distribution de probabilité jointe $=$ représentation compacte des propriétés statistiques d'un domaine
- Injecter de la connaissance *a priori* (origine : systèmes experts)
- Tirer parti de la théorie des graphes
- Connaître la probabilité d'un groupe de faits connaissant un autre groupe de faits (inférence bayésienne)

Exemple : arrosage automatique I

P(C=f)	P(C=t)
0.5	0.5



C	P(S=f)	P(S=t)
f	0.5	0.5
t	0.9	0.1

C	P(R=f)	P(R=t)
f	0.8	0.2
t	0.2	0.8

S R	P(W=f)	P(W=t)
f f	1.0	0.0
t f	0.1	0.9
f t	0.1	0.9
t t	0.01	0.99

$$P(W,S,R,C) = P(W|S,R,C).P(R|S,C).P(S|C).P(C)$$
$$P(W,S,R,C) = P(W|S,R).P(R|C).P(S|C).P(C)$$

Exemple : arrosage automatique II

Gain

- 18 paramètres sont donnés mais seulement 9 sont utiles : somme de chaque ligne $= 1$
- Sans les indépendances conditionnelles $2^4 = 16$ paramètres devraient être connus pour caractériser P(W,S,R,C)
- La structure du réseau et les tables incluent de la connaissance *a priori*

Représentation compacte

Gain en général

Pour un réseau contenant n variables booléennes, chacune pouvant être influencée par k autres variables au plus :

- Chaque CPT contient au plus 2^k valeurs
- Le réseau complet nécessite $n.2^k$ valeurs
- Sans tenir compte des dépendances : 2^n valeurs

Example

$n = 30,\ k = 5 \rightarrow$ il faut 960 valeurs contre plus d'un million

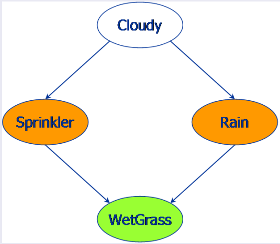
Types d'inférence

Inférence $=$ trouver $P(X|e)$

- **Causale ou Top-Down** : Calculer la probabilité d'une variable étant donné ses parents (des causes aux effets)
- **Diagnostic ou Bottom-up** : Calculer la probabilité d'une variable étant donné ses descendants (des effets aux causes)
- **Explication** : Un peu des deux
- **Reconnaissance** : Probabilité jointe complète

Diagnostic

Pourquoi la pelouse est-elle humide ?



$P(S = t|W = t) = 0.430,$
$P(R = t|W = t) = 0.708,$
$P(W = t) = 0.647$

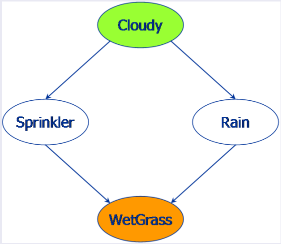
La pelouse a plus de chances d'être humide dans l'absolu. Si elle est humide c'est plus probablement parce qu'il a plu

$$P(S = t|W = t) = \frac{P(S = t, W = t)}{P(W = t)} = \frac{\sum_{c,r} P(C = c, R = r, S = t, W = t)}{\sum_{c,r,s} P(C = c, R = r, S = s, W = t)}$$

$$P(W = t) = \sum_c \sum_r \sum_s P(C = c, R = r, S = s, W = t)$$
$$P(W = t) = \sum_c \sum_r \sum_s P(C = c).P(S = s|C = c).P(R = r|C = c).P(W = t|R = r, S = s)$$
$$P(W = t) = \sum_c P(C = c). \sum_r P(R = r|C = c). \sum_s P(S = s|C = c).P(W = t|R = r, S = s)$$

Inférence Causale ou Prédiction

Si le ciel est nuageux, quel est l'état de la pelouse ?



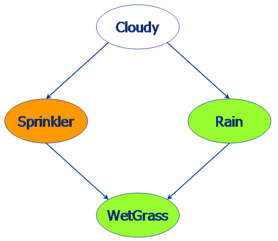
$P(W = t|C = t) = 0.7452$

Si le ciel est nuageux, il est plus probable qu'il va pleuvoir et donc que la pelouse soit humide

$$P(W = t|C = t) = \frac{P(W = t, C = t)}{P(C = t)} = \frac{\sum_{r,s} P(R = r, S = s, W = t, C = t)}{\sum_{r,s,w} P(R = r, S = s, W = w, C =)}$$

Explication

Sachant qu'il pleut et que l'herbe est humide, quid de l'état de l'arroisoir ?



$P(S = t|R = t, W = t) = 0.1945$
S'il pleut et que l'herbe est humide il y a tout de même une probabilité de 0.1945 que l'arrosoir fonctionne

Explication : inférence sur une variable de même "niveau" qu'une information.

Classification

- On se donne plusieurs réseaux de même topologie (ou pas) mais dont les probabilités sont différentes et reflètent différentes classes
- On choisit le réseaux qui fournit la plus grande probabilité jointe pour l'observation courante de toutes les variables.

d-séparation I

Définition

Deux ensembles de variables sont indépendants si ils sont d-séparés

Définition

Deux variables A et B sont d-séparés par l'ensemble E si :

- *Le chemin allant de A à B passe par un noeud d ayant des flèches convergentes et ni d ni aucun de ses descendants n'appartient à E*
- *Le chemin passe par un noeud de E mais ce noeud n'est pas convergent*

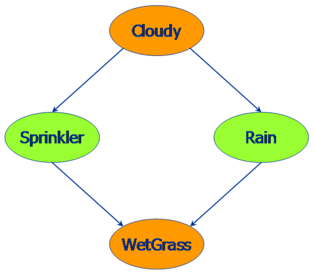
d-séparation II

Intérêt

Ces conditions indiquent comment l'information peut circuler dans le réseau
Utilité : connaître les groupes de variables indépendantes pour simplifier l'inférence (factorisation optimisée de la probabilité jointe)

d-séparation III

Example

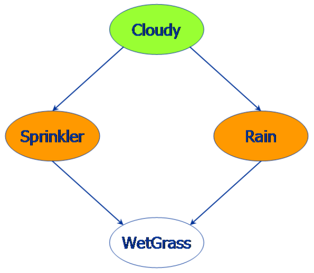


Cloudy et WetGrass sont d-séparées par $E = (Sprinkler, Rain)$ car

- Il faut passer par (S, R) pour aller de C à W
- Ni S, ni R n'ont de flèches qui convergent

d-séparation IV

Exemple

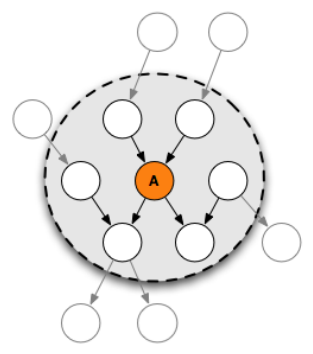


Sprinkler et Rain sont d-séparées par $E = (Cloudy)$ car

- Le chemin du haut pour aller de S à R passe par C qui n'a pas de flèches convergentes
- Le chemin du bas pour aller de S à R passe par W qui a des flèches convergentes mais n'appartient pas à E

Couverture de Markov

Définition



La couverture de Markov d'une variable A contient

- *Ses parents*
- *Ses enfants*
- *Les parents de ses enfants*

Un variable A est indépendante des autres variables du réseaux si les noeuds de sa couverture de Markov sont connus.

Inférence Exacte I

Brute Force

L'inférence exacte peut rapidement devenir suffisamment complexe à calculer pour devenir impossible en pratique.

$$P(S=t|W=t) = \frac{P(S=t,W=t)}{P(W=t)} = \frac{\sum_{c,r} P(C=c,R=r,S=t,W=t)}{\sum_{c,r,s} P(C=c,R=r,S=s,W=t)}$$

$$P(W=t) = \sum_c \sum_r \sum_s \quad P(C=c).P(S=s|C=c). \\ P(R=r|C=c).P(W=t|R=r,S=s)$$

$$P(S=t,W=t) = \sum_c \sum_r \quad P(C=c).P(S=t|C=c). \\ P(R=r|C=c).P(W=t|R=r,S=t)$$

Il faut réaliser des sommes de produits : complexité en $\mathcal{O}(n2^n)$

Inférence Exacte II

Amélioration possible

$$P(S=t|W=t) = \frac{P(S=t,W=t)}{P(W=t)} = \frac{\sum_{c,r} P(C=c,R=r,S=t,W=t)}{\sum_{c,r,s} P(C=c,R=r,S=s,W=t)}$$

$$P(W=t) = \sum_c \sum_r \sum_s \quad P(C=c).P(S=s|C=c).$$
$$P(R=r|C=c).P(W=t|R=r,S=s)$$

$$\mathbf{P(W=t)} = \quad \sum_{\mathbf{c}} \mathbf{P(C=c)}. \sum_{\mathbf{r}} \mathbf{P(R=r|C=c)}.$$
$$\sum_{\mathbf{s}} \mathbf{P(S=s|C=c).P(W=t|R=r,S=s)}$$

On sort les variables indépendantes des sommes.
Complexité encore en $\mathcal{O}(2^n)$

Inférence Exacte III

Algorithme d'élimination de variables

Problème de la méthode précédente :

$$P(W=t) = \sum_c P(C=c).\sum_r P(R=r|C=c).$$
$$\sum_s P(S=s|C=c).P(W=t|R=r,S=s)$$

Si on calcule de gauche à droite, le terme $P(S=s|C=c).P(W=t|R=r,S=s)$ est calculé 4 fois.

L'algorithme d'élimination de variables consiste à calculer les termes de gauche à droite et à les conserver en mémoire pour ne les calculer qu'une fois.

Inférence Approchée I

Modèles génératifs

Un réseau bayésien est un modèles génératifs : ils peuvent aussi générer des données en accord avec la distribution jointe qu'ils représentent.
On peut donc utiliser des méthodes dites "Monte Carlo" qui réalisent un échantillonnage de la loi jointe représentée par le réseau afin d'obtenir $P(X|e)$.

Inférence Approchée II

Algorithme d'échantillonnage direct logique

- On génère N événements en partant des "racines" jusqu'aux "feuilles"
- On recherche les événements contenant les observations e
- On rapporte le nombre de ces événements au nombre total et on obtient le résultat recherché

$$P(X|e) = \frac{(X, e)}{N}$$

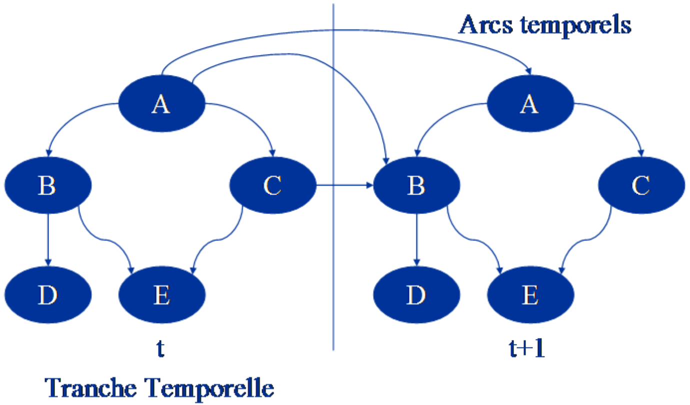
Apprentissage dans les réseaux bayésiens

Différents problèmes

- Au début, les RB étaient utilisés pour encoder la connaissance d'experts.
- Aujourd'hui on apprend sur des données

	Structure Connue	Structure Inconnue
Données Complètes	Estimation Statistique des paramètres (Max Vrais., App Bayésien)	Optimisation discrète sur les structures
Données Incomplètes	Optimisation Paramétrique (EM)	Combinaisons de méthodes (EM Structurel)

Réseaux Bayésiens Dynamiques I



Réseaux Bayésiens Dynamiques II

Souvent on suppose ...

- ... que la propriété de Markov est réalisée. Donc, on n'utilise que 2 configurations successives du réseau (pas de liens de t à t+2),
- ... que les variables sont les mêmes à tous les temps. On ne dessine pas de nouveaux noeuds d'un pas à l'autre
- ... que les liens entre variables restent inchangés à chaque étape. Donc, on redessine les mêmes arcs d'un pas à l'autre.

Réseaux Bayésiens Dynamiques III

Inférence

- Inférence exacte par déroulement (unrolling) : créer un réseau bayésien normal en "déroulant" le réseau dynamique et éventuellement appliquer l'algorithme d'élimination de variables.
- Inférence approchée par adaptation de l'algorithme de pondération par la vraisemblance.
- Débouche sur les filtres particulaires

Apprentissage

Idem BN $\rightarrow$ recherche

Applications des Réseaux Bayésiens I

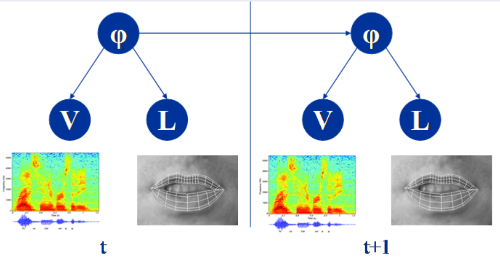
Réseaux Bayésiens

- Filtres anti-spam : la plupart des filtres aujourd'hui (Thunderbird) utilise des RB
- Mauvais payeurs : ATT a été une des premières sociétés à utiliser les RB pour la recherche dans leurs bases de données d'appels des infos pour détecter les fraudeurs (Supélec : fraude à la carte bancaire)
- Diagnostic de panne Ricoh a des RB dans ses photocopieurs
- Microsoft : l'assistant Office est un réseau bayésien, diagnostic sur les réseaux informatiques

Applications des Réseaux Bayésiens II

Réseaux Bayésiens Dynamiques

- Reconnaissance Vocale Multimodale



- Assistant Office (Aspect dynamique) : se rappelle des précédentes formulation de requêtes pour interpréter les suivantes.
- Assistance médicale : Mesure de l'évolution des paramètres physiologiques

7 Introduction

8 Réseaux Bayésiens

9 Modèles Temporels Particuliers
- Introduction
- Chaînes de Markov
- Prédiction
- Filtrage Bayésien
- Lissage
- Séquence d'états

Introduction I

Pourquoi ?

- Traitement de séquences temporelles ou spatiales d'événements difficile avec systèmes non stochastiques (ANN, SVM etc)
- Prise en compte de transformations locales et non-affines
- Apprentissage en ligne : tenir compte de nouvelles informations pour améliorer les performances

Introduction II

Exemples

- Utilisation de la dynamique dans la reconnaissance automatique de caractères manuscrits
- Suite de phonèmes en reconnaissance automatique de parole
- Séquence d'images pour la reconnaissance de gestes
- Estimation de la position d'un avion sur un radar
- Suivi de position en robotique

Introduction III

Etats et Observations

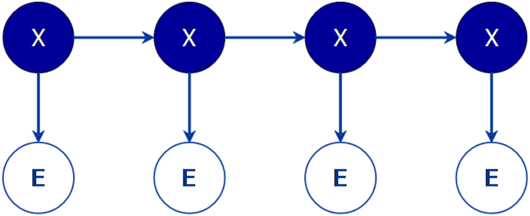
- Le monde est considéré comme partiellement observable.
- Le monde peut prendre des états x dans un ensemble X
- L'état du monde au temps t est x_t et la variable aléatoire correspondante est X_t
- L'agent dispose d'observations e issues d'un ensemble E
- L'observation e_t au temps t ne dépend que que l'état courant x_t et $P(e_t|X_t)$ est appelé modèle de capteur ou d'observation.
- Les observations arrivent à partir du temps $t_0 + 1$ (on prend souvnet $t_0 = 0$)

Introduction IV

Stationnarité et propriété de Markov

- Le processus étudié est stationnaire : $P(X_t|e_t)$ et $P(X_t|X_{0:t-1})$ ne dépendent pas de t
- On suppose que le processus est Markovien (d'ordre 1) : $P(X_t|X_{t-1})$

Remarque : On peut toujours se ramener à un ordre 1 par ajout de variables d'état.



Introduction V

Paramètres du modèle

- Probabilités de transitions $P(X_t|X_{t-1})$
- Modèle de capteur : $P(E_t|X_t)$
- Distribution de probabilité *a priori* sur les états initiaux $P(X_0)$

Evolution du monde

$$P(X_0, X_1, X_2, ..., E_1, E_2, ...) = P(X_0)\prod_{i=1}^{t} P(X_i|X_{i-1})P(E_i|X_i)$$

Introduction VI

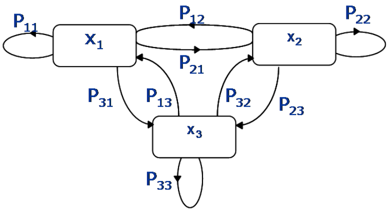
5 types de problèmes

- **Prédiction** : quel est l'état du monde le plus probable au **temps futur** $t + k$ étant donné une séquence d'observations ($P(X_{t+k}|e_{1:t})$)
- **Filtrage** (monitorage) : quel est l'état du monde le plus probable au **temps courant** t étant donné une séquence d'observations ($P(X_t|e_{1:t})$)
- **Lissage** (rétrospection) : quel est l'état du monde le plus probable au **temps passé** $t - k$ étant donné une séquence d'observations ($P(X_{t-k}|e_1 : t)$)
- **Explication** : quelle est la séquence d'état du monde la plus probable étant donné une séquence d'observations ($P(X_{0:t}|e_{1:t})$)
- **Entraînement** : apprendre les paramètres du modèle

Cas simple : Chaînes de Markov I

Définition

- *Automate à états finis stochastiques : monde totalement observable*
- *Défini par sa topologie (états X_l, transitions), et par des probabilités de transitions (P_{ij}).*
- *Entraînement : comptage (maximum de vraisemblance)*



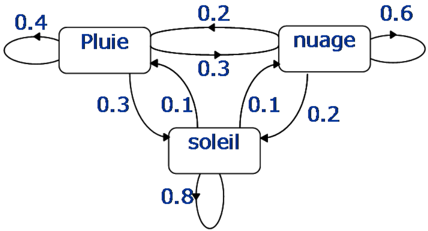
Définition

- x_l *est le* l^{eme} *état de la chaîne*
- x^n *est l'état visité au temps n*
- x_l^n *est l'événement* $x_l = x^n$
- $P_{ij} = P(x_j^n | x_i^{n-1})$

Cas simple : Chaînes de Markov II

Exemples d'utilisation

On cherche : $P(S-S-S-R-R-S-C-S) = P(S).P(S|S).P(S|S).P(R|S).P(R|R)... = (1.0)(0.8)(0.8)(0.1)(0.4)...$



Cas simple : Chaînes de Markov III

Quelques opérations utiles

- Probabilité d'un chemin : $P(x^1, x^2, ..., x^N) = \prod_{n=1}^{N-1} P(x^n|x^{n-1})$
- Probabilité de rester dans l'état x_i pendant exactement d pas $P(x_i^0, x_i^1, x_i^2, ..., x_i^d, x_j^{d+1}) = (P_{ii})^d(1 - P_{ii})$
- Probabilité d'aller de l'état x_ià l'état $x_j$en exactement N pas : $P(x_j^N|x_0^i) = \sum_{n=0}^{N}\sum_{l=1}^{L} P(x_j^N, x_l^n|x_i^0)$
- Probabilité du meilleurs chemin à N pas entre 2 états $P_{n+1}^{best}(l) = \max_k P_n^{best}(k)P_{kl}$

Prédiction

Problème : prédiction à k pas

Estimer la probabilité *a posteriori* $P(X_{t+k+1}|e_{1:t})$ de l'état du monde au temps futur $t+k+1$ étant donné la séquence d'observations $e_{1:t}$

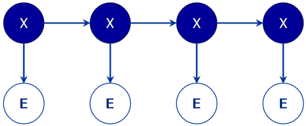
Solution formelle

$$p(X_{t+k+1}|E_{1:t}) = \int_X p(X_{t+k+1}|X_{t+k})p(X_{t+k}|e_{1:t})dX_{t+k}$$

Récursivité avant (forward)

La prédiction est donc un calcul récursif partant de t jusqu'à $t+k$: on calcule $p(X_{t+k+1}|E_{1:t})$ grâce à $p(X_{t+k}|e_{1:t})$

Filtrage Bayésien I



Position du problème

$$x_{k+1} = f_k(x_k, v_k) \text{ (evolution)}$$
$$e_k = g_k(x_k, n_k) \text{ (observation)}$$

Solution formelle : Récursivité avant

$$p(X_{k-1}|E_{1:k-1}) \xrightarrow[p(X_k|X_{k-1})]{\text{prédiction}} p(X_k|E_{1:k-1}) \xrightarrow[p(E_k|X_k)]{\text{correction}} p(X_k|E_{1:k})$$

Prédiction :$p(X_k|E_{1:k-1}) = \int_{\mathcal{X}} p(X_k|X_{k-1})p(X_{k-1}|E_{1:k-1})dX_{k-1}$

Correction :$p(X_k|E_{1:k}) = \frac{p(E_k|X_k)p(X_k|E_{1:k-1})}{p(E_k|E_{1:k-1})}$

avec $\quad p(E_k|E_{1:k-1}) = \int_{\mathcal{X}} p(E_k|X_k)p(X_k|E_{1:k-1})dX_k$

Filtre de Kalman I

Position du problème

On dispose de $e_{0:k} = \{e_0, e_1, ..., e_k\}$
Le système est supposé linéaire (et stationnaire) :

$$\begin{aligned} X_k &= AX_{k-1} + BU_k + V_k \\ e_k &= CX_k + DU_k + W_k \end{aligned}$$

Les bruits sont supposés blancs, gaussiens, centrés et décorrélés

$$\begin{aligned} V_k &= \mathcal{N}(0, \Sigma_V) \\ W_k &= \mathcal{N}(0, \Sigma_W) \\ e[v \cdot w] &= 0 \end{aligned}$$

Filtre de Kalman II

Propriétés importantes de l'algorithme

- Fonctions d'évolution et d'observation linéaires
- Bruits gaussiens
- Propagation de statistiques suffisantes de l'état caché à travers les transformations linéaires.
- Famille gaussienne fermée pour les transformations linéaires
- Algorithme itératif

Filtre de Kalman III

On cherche le vecteur moyen $X_{k:k}$ (estimation de l'état) et la matrice de covariance $P_{k:k}$(estimation de l'erreur)

$$\widehat{X_{k:k}} = E[X_k | e_{0:k}]$$

On veux minimiser l'énergie de l'erreur :

$$E[(X_k - \widehat{X_{k:k}})^T (X_k - \widehat{X_{k:k}})]$$

C'est aussi la trace de la matrice de covariance de l'erreur :

$$trace(P_{k:k}) = trace(E[(X_k - \widehat{X_{k:k}})(X_k - \widehat{X_{k:k}})^T])$$

On veut que l'algorithme soit itératif !

Filtre de Kalman IV

Etape de prédiction

$$\begin{aligned}
\widehat{X_{k:k-1}} &= E[X_k | e_{0:k-1}] \\
\widehat{X_{k:k-1}} &= E[AX_{k-1} + BU_k + V_k | e_{0:k-1}] \\
\widehat{X_{k:k-1}} &= A\widehat{X_{k-1:k-1}} + BU_k
\end{aligned}$$

Filtre de Kalman V

Etape de correction

$$P_{k:k-1} = AP_{k-1:k-1}A^T + \Sigma_V$$

$$\begin{aligned} \widehat{X_{k:k}} &= \widehat{X_{k:k-1}} - K_k[e_k - (C\widehat{X_{k:k-1}} + DU_k)] \\ P_{k:k} &= [I - K_k C]P_{k:k-1} \end{aligned}$$

Avec K_k, le gain optimal de Kalman :

$$K_k = P_{k:k-1}C^T[CP_{k:k-1}C^T + \Sigma_W]^{-1}$$

Remarque : les $P_{k:k}$ et K_k sont indépendant des observations et peuvent donc être calculés à l'avance !

Cas non-linéaire

Filtre de Kalman étendu (EKF)

- Extension au cas des fonctions non-linéaires
- Bruits toujours gaussiens
- Linéarisation des fonctions autour de l'état estimé

Filtre de Kalman non-parfumé (UKF)

- Fonctions non-linéaires
- Génération de points à partir d'estimation de distribution de l'état caché
- Passage de ces points à travers les fonctions non-linéaires
- Nouveau calcul des statistiques suffisantes à partir de ces nouveaux points

Cas non-gaussien

Filtrage particulaire

- Version tout numérique du filtrage bayésien
- On approxime les probabilités a posteriori par des sommes de Dirac
- Appelés aussi méthodes de Monte Carlo séquentielles
- Plus de calcul de statistique suffisante

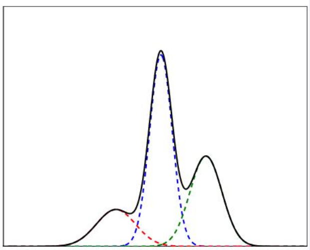
Example : Regression I

Gaussian Mixture

Approximation of a function $f(x)$ by a mixture of Gaussians $\hat{f}_\theta(x)$

$$\hat{f}_\theta(x) = \sum_{i=1}^{p} \omega_i e^{\frac{x-\mu_i}{\sigma_i^2}}$$

where $\theta = [(\omega_i, \alpha_i, x_i)]^T$



- We can search for the best ω_i but also μ_i and σ_i .
- note : priors θ_0 and $P_{0|0}$ have to be chosen

Example : Regression II

State-space representation

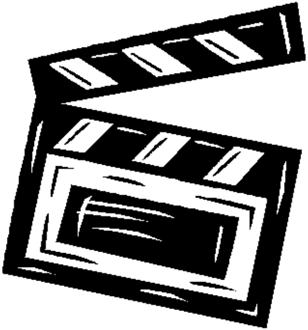
$$\theta_k = \theta_{k-1} + V_k \text{ (Evolution equation)}$$

$$e_k = \sum_{i=1}^{p} \omega_i e^{\frac{x-\mu_i}{\sigma_i^2}} + W_k \text{ (observation equation)}$$

Experiments

- Noisy Cardinal Sine

$$f(x) = \frac{sin(x)}{x}, \quad x \in [-10, 10]$$



Method	test error	# kernels
Proposed algorithm	**0.0385 (-25.8%)**	**9.6 (-65.7%)**
Figueiredo	0.0455 (-12.3%)	7.0 (-75%)
SVM	**0.0519 (-0.0%)**	**28.0 (-0%)**
RVM	0.0494 (-4.8%)	6.9 (-75.3%)
VRVM	0.0494 (-4.8%)	7.4 (-73.5%)
MCMC	0.0468 (-9.83%)	6.5 (-76.8%)
Sequential RVM	0.0591 (+13.8%)	4.5 (-83.9%)

Example 2 : non linear observation equation I

Problem

- It is not necessary to directly observe the function to regress
- Let's try to regress f while observing a non-linear mapping $g_f(x, t(x))$ where the non-linear function g is known but t is only observed
- We search for an approximation of f_θ given observations

$$(x_k, t_k = t(x_k), y_k = g_f(x_k, t_k) + n_k)_k$$

- State-space representation :

$$\begin{cases} \theta_{k+1} = \theta_k + v_k \\ y_k = g_{\hat{f}_{\theta_k}}(x_k, t_k) + n_k \end{cases}$$

Example 2 : non linear observation equation II

Setting

- Let f be a 2D cardinal sine en 2D on $[-10, 10]^2$

$$f(x) = f(x^1, x^2) = \frac{sin(x^1)}{x^1} + \frac{x^1 x^2}{100}$$

- Let the observed non-linear mapping t be

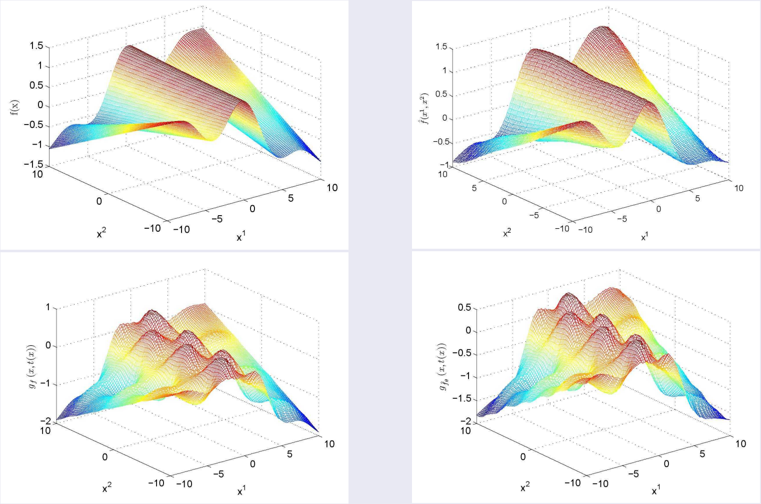
$$t(x) = t(x^1, x^2) = 10 \tanh(\frac{x^1}{7}) + \sin(x^2)$$

- Let the known non linear function g be

$$g_f(x, t(x)) = f(x) - \gamma \max_{z \in [-10,10]} f(t(x), z)$$

Example 2 : non linear observation equation III

Results

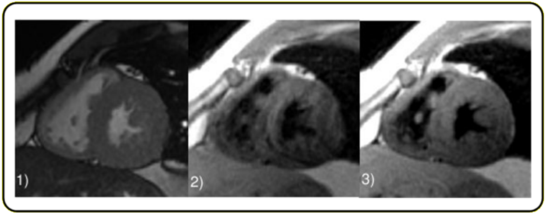


Cardiac IRM acquisition

Problem : one image = several acquisitions but the heart moves
Prediction of cardiac phases by measuring RR intervals (Δ), breathing (S), ECG magnitude (A)

$$\Delta_{n+1} = \sum_{i=0}^{n_1} a_i \Delta_{n-i} + \sum_{j=0}^{n_2} b_j S_{n-j} + \sum_{k=0}^{n_3} c_k dS_{n-k} + \sum_{l=0}^{n_4} d_l A_{n-l}.$$

$$\begin{cases} \theta_k = \theta_{k-1} + w_{k-1} \\ \Delta_k = C_k \theta_k + v_k \end{cases},$$



Lissage I

Position du problème

On cherche la distribution sur les états passés étant donné une séquence d'observation : $P(X_k|e_{0:t})$ avec $1 \leq k < t$

Solution formelle

$$
\begin{aligned}
P(X_k|e_{1:t}) &= P(X_k|e_{1:k}, e_{k+1:t}) \\
&= \alpha P(X_k|e_{1:k}) P(e_{k+1:t}|X_k, e_{1:k}) (Bayes) \\
&= \alpha \underbrace{P(X_k|e_{1:k})}_{Prediction, forward} P(e_{k+1:t}|X_k)(ind.cond.)
\end{aligned}
$$

Lissage II

Calcul récursif arrière (Backward)

$$
\begin{aligned}
P(e_{k+1:t}|X_k) &= \int_{x_{k+1}} P(e_{k+1:t}|X_k, xk+1)P(x_{k+1}|X_k) \\
&= \int_{x_{k+1}} P(e_{k+1:t}|xk+1)P(x_{k+1}|X_k) \\
&= \int_{x_{k+1}} P(e_{k+1}, e_{k+2:t}|xk+1)P(x_{k+1}|X_k) \\
&= \int_{x_{k+1}} P(e_{k+1}|, xk+1)P(e_{k+2:t}|xk+1)P(x_{k+1}|X_k)
\end{aligned}
$$

Lissage III

Algorithme Forward-Bacward

$$P(X_k|e_{1:t}) = \alpha \underbrace{P(X_k|e_{1:k})}_{\textit{Forward}} \underbrace{P(e_{k+1:t}|X_k)}_{\textit{Backward}}$$

Séquence d'états ayant généré les observations I

Position du problème

On cherche la séquence d'états $(x_1, x_2, \ldots, x_t)$ ayant le plus vraisemblablement généré la séquence d'observations $(e_1, e_2, \ldots, e_t)$

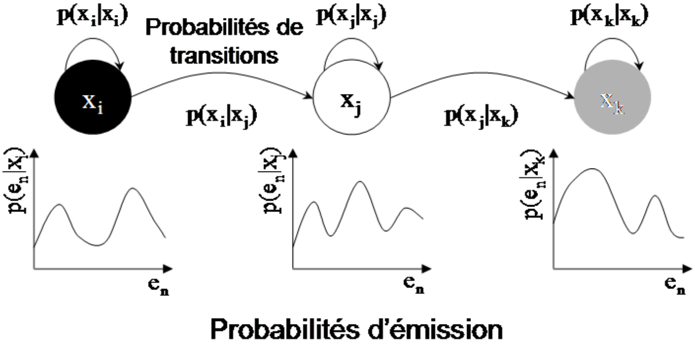
Algorithme de Viterbi

$$\max_{x_1,\ldots,x_t} P(x_1, \ldots, x_t, X_{t+1}|e_{1:t}) =$$

$$\alpha P(e_{1:t}|X_{t+1}) \max_{x_t}(P(X_{t+1}|x_t) \max_{x_1,\ldots,x_{t-1}} P(x_1, \ldots, x_{t-1}, x_t|e_{1:t}))$$

Le chemin optimal est composé de sous-chemins optimaux (programmation dynamique, Bellman)

Modèles de Markov Cachés I

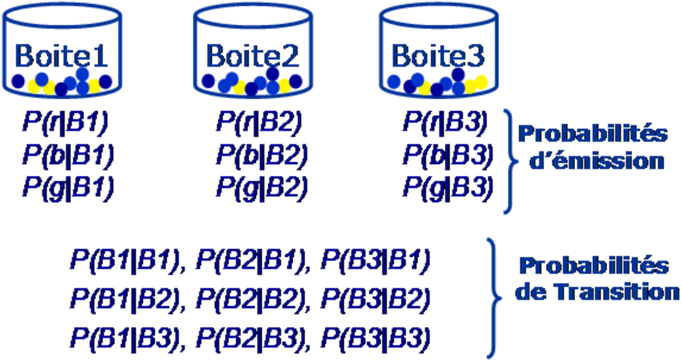


Processus doublement stochastique : transitions et émissions

Peut se mettre sous forme matricielle !

Modèles de Markov Cachés II

Exemple :



Modèles de Markov Cachés III

3 problèmes :

Estimation : $P(e|M)$

Trouver la probabilité d'une séquence d'observation étant donné le modèle.

Entraînement

Trouver les probabilités de transitions et d'émissions étant donné une topologie et des données

Décodage

Séquence d'état la plus probable étant donné une séquence d'observations

Modèles de Markov Cachés IV

3 solutions :

Estimation

- Complexité : $\mathcal{O}(N^{observations})$
- 2 algorithmes : Baum Welch (tous les chemins) et Viterbi (meilleur chemin)

Entraînement

- Souvent on n'a que des données non étiquetées
- Algorithme EM

Décodage

Algorithme de Viterbi

Modèles de Markov Cachés V

Applications

- Reconnaissance vocale
- Reconnaissance de gestes
- Reconnaissance d'écriture/signature

Sites intéressants

- HTK : http://htk.eng.cam.ac.uk
- OpenCV : http://sourceforge.net/projects/opencvlibrary
- Torch : http://www.torch.ch

Troisième partie III

Modèles Graphiques pour le Contrôle

Types de décision I

Décision de test

A partir d'observations générées par un environnement, l'agent prend une décision sur l'état de cet environnement.

Exemples

- Classification
- Détection
- Diagnostic
- Inférence dans les réseaux bayésiens.

Types de décision II

Décision d'action

A partir d'une succession d'actions et d'observations, l'agent prend une décision sur l'action à accomplir à l'instant présent. Cette action aura le plus souvent pour effet de modifier l'environnement et de générer une nouvelle observation.

Exemples

- Contrôle optimal
- Jeux

Types de décision III

En pratique ...

... ce n'est pas si simple ! Chaque décision à une composante de test (identification de l'état) et une composante d'action (en tirer les conséquences)

Exemples

- Diagnostic suivi d'un traitement
- Observation topologique en vision par ordinateur et action d'un robot

Position du problème général

Etats, actions, observations

Un **agent** est immergé dans un **environnement**. Il reçoit des **observations** E générées par cet environnement, qu'on appelle aussi **perceptions**. En utilisant ces perceptions, l'agent infère un **état** S probable de l'environnement (degré de croyance) et il doit réaliser une **action** A qui va l'amener vers un état $S_i(A, S)$ plus désirable dans l'environnement (désir, volonté, besoin).

Action

Une action est donc une fonction de l'état S $(A(S))$. A chaque action, plusieurs états de sortie $S_i(A, S)$ sont possibles avec chacun une probabilité $P(S_i|A, S)$.

Théorie de la décision I

Rappel :

Dans le cas de la décision de test, nous avons déjà défini l'utilité associée à une décision

$$u(h|f),$$

l'utilité moyenne associée à la décision d'associer l'événement e à l'hypothèse h

$$U(h|e) = \sum_f u(h|f)P(f|e),$$

et la décision optimale :

$$h^* = \underset{h}{\operatorname{argmax}}\, U(h|e) \;\; si \max_h U(h|e) > \sigma_U$$

Théorie de la décision II

Utilité

Ici, l'utilité sera associée à un état : $U(S)$. C'est un nombre réel qui détermine dans quelle mesure il est bon pour l'agent d'être dans l'état S.

L'utilité moyenne d'une action sachant les observations E est alors donnée par :

$$E[U(A|E)] = \sum_i P(S_i(A,S)|A,E)U(S_i(A,S))$$

La décision optimale est celle qui maximise l'utilité moyenne : règle MEU (Maximum Expected Utility) :

$$A^* = \underset{A}{\operatorname{argmax}}\, E[U(A|E)]$$

10 Introduction

11 Réseaux de Décisions
- Définition
- Exemple
- Types d'action
- Valeur de l'information
- Dynamique
- Remarque et Conclusions

12 Markov Decision Processes

13 Programmation Dynamique

14 Apprentissage par Renforcement

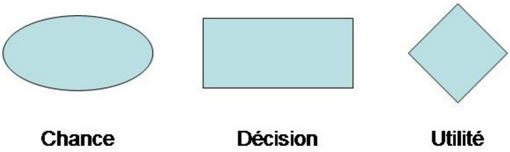
Définition I

Réseau de décisions

Un réseau de décisions est une extension des réseaux bayésiens dans lesquels ont été ajoutés différents types de noeuds :

- Nœuds de chance : de forme ovale, ils représentent des variables aléatoires décrivant l'état de l'environnement, comme dans les réseaux bayésiens standards.
- Nœuds de décisions : de forme rectangulaire, ils représentent la décision prise à un instant donné. Leurs valeurs sont les différentes actions possibles.
- Nœuds d'utilité ou de valeur : en forme de losange, ils représentent l'utilité de l'action choisie étant donné l'état du système.

Définition II

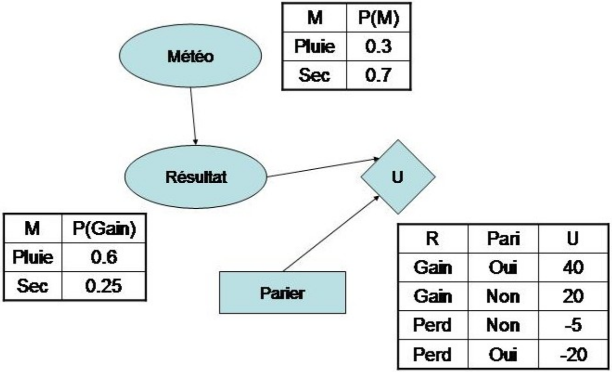


Remarques

- *Les parents d'un nœud de décision peuvent être des nœuds de chance, signifiant que ces variables doivent être connues pour prendre la décision (lien d'information).*
- *Les parents d'un nœud d'utilité sont les variables aléatoires qui décrivent l'état du système. On peut donc avoir des nœuds de décision.*
- *On les appelle aussi des réseaux d'influence.*

Exemple I

Pari sur un match de football



M	P(M)
Pluie	0.3
Sec	0.7

M	P(Gain)
Pluie	0.6
Sec	0.25

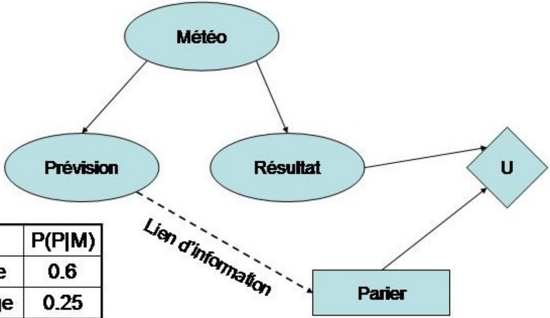
R	Pari	U
Gain	Oui	40
Gain	Non	20
Perd	Non	-5
Perd	Oui	-20

$$EU(Paris = Oui) = P(R = gain)U(R = gain|Paris = Oui) + P(R = perd)U(R = perd|Paris = Oui) = 1.3$$

$$\mathbf{EU(Paris = Non) = 3.875}$$

Exemple II

Pari avec lien d'information



M	P	P(P|M)
Pluie	Pluie	0.6
	Nuage	0.25
	Soleil	0.15
Sec	Pluie	0.10
	Nuage	0.4
	Soleil	0.5

P	Pari
Pluie	Oui
Nuage	Non
Soleil	Non

Algorithme de résolution

`Algorithme de prise de décision`

Insérer les observations disponibles

Pour chaque combinaison possible des valeurs des parents du noeud de décision **faire**

    Pour chaque action disponible **faire**

1. Utiliser la valeur de l'action courante
2. Calculer les probabilités a posteriori des parents du noeud d'utilité
3. Calculer l'utilité moyenne pour l'action

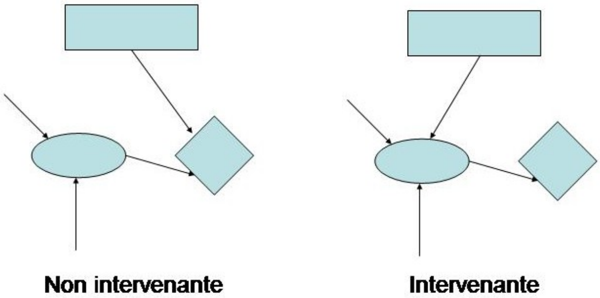
    fin pour

    enregistrer l'action avec la plus grande utilité

fin pour

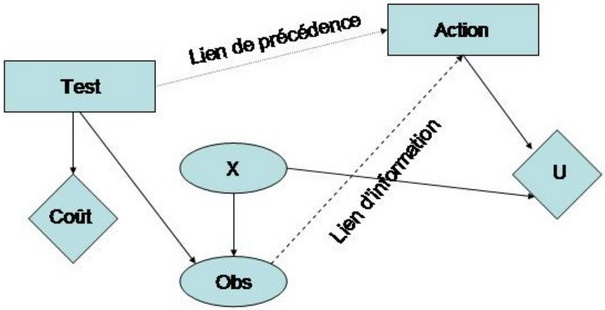
retourner la table de décision

Types d'action



- Action non-intervenante : pas de conséquence sur le monde (pari)
- Action intervenante : action directe sur le monde (prescription médicale)

Décisions séquentielles



Les DN ne se prêtent pas bien à la décision séquentielle sauf dans le cas du test.

Valeur de l'information

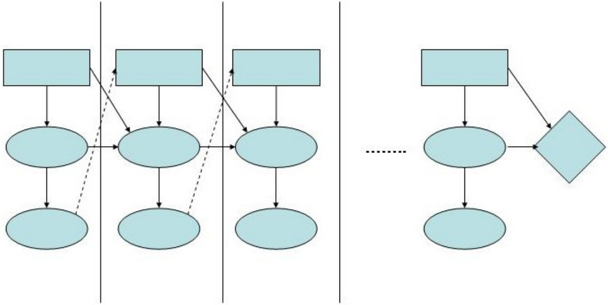
De manière générale, se pose la question de chercher à obtenir des informations pour prendre une décision la plus utile possible. Mais les actions permettant de recevoir plus d'information peuvent coûter et donc diminuer l'utilité globale. On associe donc une valeur à l'information.

Dans le cas simple du test :

$$V(test) = EU(Test = oui) - EU(Test = non)$$

C'est le *prix* maximum qu'on peut être prêt à payer pour obtenir l'information.

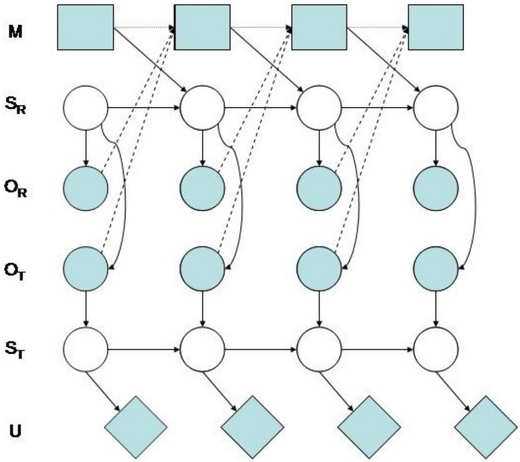
Réseau de décisions dynamique I



Remarque : il y a un seul noeud d'utilité, en fin d'interaction. Ce n'est pas toujours le cas.

Réseau de décisions dynamique II

Exemple : robot suivant une cible



Remarque et conclusions

Remarque

Les arbres de décisions peuvent aussi être modélisés sous forme de réseau de décision.

Conclusions

- Les réseaux de décisions permettent la modélisation du choix simple d'une action,
- Ils se basent sur le principe MEU (Maximum Expected Utility)
- Mêmes principes d'inférence que les réseaux bayésiens,
- Ils s'étendent à la décision séquentielle pour le cas du test en suivant le principe de la valeur de l'information,
- Ils s'étendent au cas dynamique.

Types d'apprentissage

Types d'apprentissage

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

Types d'apprentissage

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

Apprentissage Non-Supervisé

- Apprendre une structure dans un ensemble de données ;
- Pas d'oracle ;

Types d'apprentissage

Apprentissage Supervisé

- Apprendre des relations entre entrées et sorties ;
- Un oracle donne des exemples exprimant ces relations ;

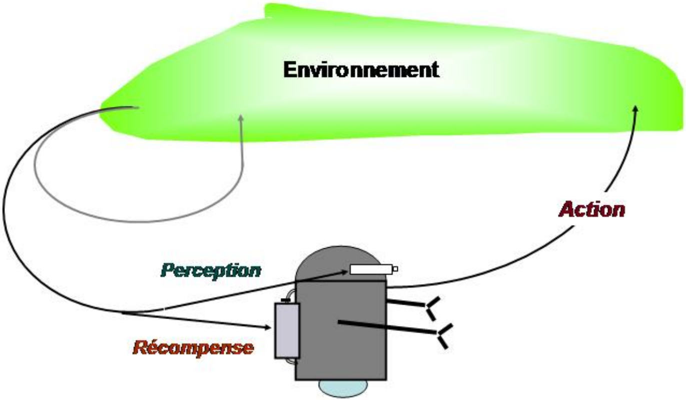
Apprentissage Non-Supervisé

- Apprendre une structure dans un ensemble de données ;
- Pas d'oracle ;

Apprentissage par Renforcement

- Apprendre à se comporter !
- Apprentissage en ligne
- Décisions séquentielles

Problème général



Problèmes induits

Apprentissage par Essais-Erreurs

- L'apprenant est obligé d'agir pour apprendre.

Dilemme Exploration vs Exploitation

- Doit-on continuer à suivre une stratégie dont on connaît les conséquences ?
- Doit-on explorer l'environnement pour trouver une meilleure stratégie ?

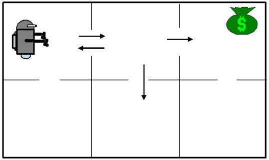
Récompenses Retardées

- Les fruits d'une action peuvent-être retardés
- Comment sacrifier des petits gains à court terme pour privilégier les gains maximum à long terme

Exemples

Jouets

- Robot dans un labyrinthe
- Pendule inversé
- n-armed bandit
- Echecs, Go, Backgammon ...



Réels

- Exploration Spatiale (Robot sur Mars)
- Secours dans les décombres
- Comportement en milieu hostile et/ou inconnu en général

Historique I

Psychologie animale - 1900

- Recherches initiées par Pavlov et Skinner début XXème
- 1911 : E. Thorndike, "The law of effects"

> *Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur ; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.*

Historique II

Contrôle Optimal - 1950

- 1957 - R. Bellman : Dynamic Programming
- 1960 - R. Howard : Dynamic Programming and Markov Processes
- 1965 - K. Astrom : Optimal control of Markov decision processes with incomplete state estimation

Apprentissage Numérique - 1980

- 1978 - R. Sutton : théorie de l'apprentissage animal
- 1983 - A. Barto et R. Sutton : résolution du problème du pendule inversé par voie neuronale.
- 1989 - C. Watkins : Q-learning
- 1992 - G. Tesauro : TD-Gammon

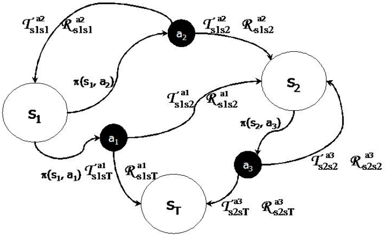
Processus Décisionnels de Markov (MDP)

Définition (PDM)

Un PDM est constitué des éléments suivants :

- *S l'espace des états ;*
- *A l'espace des actions ;*
- *T l'axe temporel ;*
- *$G \subset S \times A$ le graphe reliant chaque état à différentes actions ;*
- *$\mathcal{T}^a_{ss'} \in (p_t)_{t \in T}$ une famille de distributions de probabilité de transitions markoviennes entre actions et états*
- *$(r_t)_{t \in T}$ une famille bornée de fonctions de récompenses sur les transitions d'état.*

Notons que $S \cap A = \emptyset$ et $S \cup A \neq \emptyset$.



interprétation

À chaque instant t de T, l'agent observe l'état courant $s_t \in S$, applique une action $a_t \in A$ sur le système qui le mène aléatoirement selon $\mathcal{T}^a_{ss'} = p_t(.|s_t, a_t)$ dans le nouvel état s_{t+1} (l'expression $p_t(s'|s, a)$ représentant la probabilité de passer dans l'état s' après avoir exécuté l'action a à la date t dans l'état s), et reçoit une récompense $r_t(s_t, a_t, s_{t+1}) \in \mathbb{R}$. avec $\mathcal{R}^a_{ss'} = E[r_t|s, s', a]$

Remarques

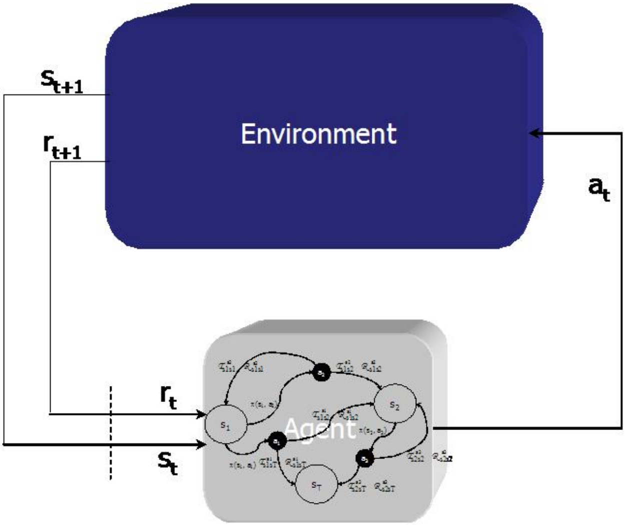
Si le choix de l'action est déterministe

- $P(s_{t+1}|s_t, a_t) = P(s_{t+1}|s_t)$
- Le MDP devient une chaîne de Markov
- Un MDP est une chaîne de Markov *contrôlée*

MDP = Modèle de l'environnement

- Le MDP est la modélisation interne du monde faite par l'apprenant
- Dans le cas parfait, l'environnement réagit de la même manière que le MDP !
- Sinon, erreur de modélisation $\Rightarrow$ incertitude !

Vue de l'Apprenant



Gain : début de la vision à long terme

Définition (Gain Cumulé)

$$R_t = r_{t+1} + r_{t+2} + \ldots + r_T = \sum_{i=t+1}^{T} r_i$$

Définition (Gain Cumulé avec Intérêt ou γ-pondéré)

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \ldots + \gamma^{T-t+1} r_T + \ldots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

Définition (Gain Moyen)

$$R_t = \frac{1}{T-1} \sum_{i=t+1}^{T} r_i$$

Stratégie

Définition (Politique ou Stratégie π)

La politique ou stratégie π_t de l'apprenant à l'instant t est une application de S (éventuellement S^n) dans A qui définit le comportement de l'apprenant (mapping entre situation et action)

politique π_t	*déterministe*	*aléatoire*
markovienne	$s_t \rightarrow a_t$	$s_t \rightarrow q_{\pi_t}(s_t)$
histoire-dépendante	$h_t \rightarrow a_t$	$h_t \rightarrow q_{\pi_t}(h_t)$

Définition (Politique ou Stratégie Optimale π^*)

La politique ou stratégie optimale π^* pour un MDP donné est une politique qui maximise le gain de l'apprenant

Fonctions de Valeur I

Il faut disposer d'une fonction d'évaluation locale du gain à long terme espéré en suivant une politique

Définition (Fonction de Valeur d'état $V^\pi(s)$)

$$\forall s \in S \quad V^\pi_\gamma(s) = E^\pi[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1}(s_t, a_t) | s_0 = s]$$

$V^\pi(s) =$ *Gain moyen espéré en partant de s et en suivant la politique* π

Apprentissage par Renforcement

Apprendre cette fonction par interaction pour en tirer des conclusions sur la politique à suivre

Fonctions de Valeur II

On peut aussi disposer d'une fonction d'évaluation locale du gain à long terme espéré en choisissant une action particulière

Définition (Fonction de Valeur d'action $Q^\pi(s,a)$)

$$\forall s \in S, a \in A \quad Q^\pi_\gamma(s,a) = E^\pi[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1}(s_t, a_t) | s_0 = s, a_0 = a]$$

$Q^\pi(s,a) =$ *Gain moyen espéré en partant de* s *et en choisissant* a *puis en suivant la politique* π

Fonctions de Valeur III

Remarque

$$Q^{\pi}_{\gamma}(s,a) = E^{\pi}[r_t(s,a) + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2}(s_t, a_t) | s_0 = s, a_0 = a]$$

$$Q^{\pi}_{\gamma}(s,a) = \sum_{s'} \mathcal{T}^{a}_{ss'}[\mathcal{R}^{a}_{ss'} + \gamma V^{\pi}(s')]$$

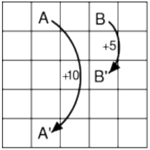
$$\forall s \in S \quad V^{\pi}_{\gamma}(s) = \sum_{a} \pi(s,a) Q^{\pi}_{\gamma}(s,a)$$

Evaluation de $V^{\pi}(s)$

$$\forall s \in S \quad V^{\pi}_{\gamma}(s) = \sum_{a} \pi(s,a) Q^{\pi}_{\gamma}(s,a)$$

$$\forall s \in S \quad V^{\pi}_{\gamma}(s) = \sum_{a} \pi(s,a) \sum_{s'} \mathcal{T}^{a}_{ss'} [\mathcal{R}^{a}_{ss'} + \gamma V^{\pi}(s')]$$

Systèmes de $|S|$ équations linéaires à $|S|$ inconnues (sauvegarde tabulaire).



(a)

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

(b)

Programmation Dynamique I

Définition (Politique Optimale π^*)

La politique optimale π^* est la politique qui maximise $V^\pi(s)$ pour tout les états. La fonction de valeur associée est notée V^*

Théorème (Équations de Bellman pour $V^*(s)$)

$$\begin{aligned} V^*(s) &= \max_\pi V^\pi(s) = \max_{a \in A} Q^*(s,a) \\ &= \max_a \sum_{s'} \mathcal{T}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma V^*(s')] \end{aligned}$$

Programmation Dynamique II

Théorème (Équations de Bellman pour $Q^*(s,a)$)

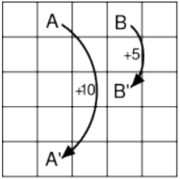
$$\begin{aligned} Q^*(s,a) &= \max_{\pi} Q^{\pi}(s,a) \\ &= \sum_{s'} \mathcal{T}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma \max_{a'} Q^*(s',a')] \end{aligned}$$

Si on connaît la dynamique du système (probabilités de transition et récompense immédiate), la politique optimale s'en déduit aisément :

$$\forall s \in S \quad \pi^*(s) = \operatorname*{argmax}_a \sum_{s'} \mathcal{T}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma V^*(s')]$$

Programmation Dynamique III

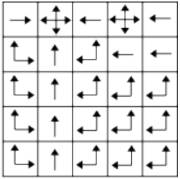
Exemple : Grid World



a) gridworld

22.0	24.4	22.0	19.4	17.5
19.8	22.0	19.8	17.8	16.0
17.8	19.8	17.8	16.0	14.4
16.0	17.8	16.0	14.4	13.0
14.4	16.0	14.4	13.0	11.7

b) V^*



c) π^*

Itération de la valeur (Value Iteration) I

MOCAD - DI

Olivier Pietquin

Introduction

Réseaux de Décisions

Markov Decision Processes

Programmation Dynamique

Politique Optimale

Value Iteration

Policy Iteration

Asynchronous DP

Apprentissage par Renforcement

POMDP

Problème avec les équations de Bellman

Système de $|S|$ equations non-linéaires (max) à $|S|$ inconnues

Théorème (Contraction)

L'équation de Bellman pour $V^*(s)$ défini un opérateur B qui est une contraction et admet un point fixe en $V^*(s)$

$$
\begin{aligned}
V_{i+1} &\leftarrow BV_i \\
V_{i+1}(s) &\leftarrow \max_a \sum_{s'} \mathcal{T}^a_{ss'} [\mathcal{R}^a_{ss'} + \gamma V_i(s')]
\end{aligned}
$$

Remarque : Une contraction est un opérateur qui, appliqué à 2 arguments différents, donne des valeurs plus proches l'une de l'autre que les arguments ne l'étaient

Itération de la valeur (Value Iteration) II

Algorithme d'itération de la valeur

initialiser $V_0 \in \mathcal{V}$

$n \leftarrow 0$

tant que $\|V_{n+1} - V_n\| > \varepsilon$ **faire**

  pour $s \in S$ **faire**

    $V_{n+1}(s) = \max_a \sum_{s'} \mathcal{T}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma V_n(s')]$

  fin pour

  $n \leftarrow n + 1$

fin tant que

pour $s \in S$ **faire**

  $\pi(s) = \text{argmax}_{a \in A} \sum_{s'} \mathcal{T}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma V_n(s')]$

fin pour

retourner V_n, π

Complexité : $\mathcal{O}(|A||S|^2)$

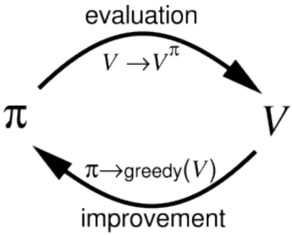
Itération de la politique (Policy Iteration) I

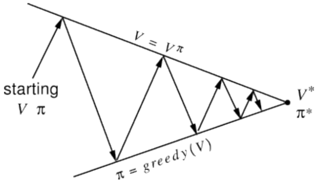
Problème avec l'itération de la valeur

Suivant la valeur d'ϵ, la convergence peut-être lente.

Amélioration

Même si on n'atteint pas l'estimation exacte de $V^*(s)$, on peut trouver la politique optimale. On va donc itérer sur la politique.





Itération de la politique (Policy Iteration) II


```
Algorithme d'itération de la politique
```

initialiser $\pi_0 \in \mathcal{D}$

$n \leftarrow 0$

tant que $\pi_{n+1} \neq \pi_n$ **faire**

résoudre (Phase d'évaluation)

$V_{n+1}(s) = \sum_{s'} \mathcal{T}_{ss'}^{\pi(s)} [\mathcal{R}_{ss'}^{a} + \gamma V_n(s')]$ (eq. Linéaire)

pour $s \in S$ **faire** (Phase d'amélioration)

$\pi_{n+1}(s) = \text{argmax}_{a \in A} \sum_{s'} \mathcal{T}_{ss'}^{a} [\mathcal{R}_{ss'}^{a} + \gamma V_n(s')]$

fin pour

$n \leftarrow n + 1$

fin tant que

retourner V_n, π_{n+1}

Complexité : $\mathcal{O}(|A||S|^2) + \mathcal{O}(|S|^3)$

Programmation Dynamique Asynchrone

Problèmes avec DP

Pour de très grands espaces d'états, le temps d'une itération peut être très important

Programmation Dynamique Asynchrone

- Mise à jour pour un état sur k
- Introduction de connaissance *a priori* dans le choix des états à mettre à jour.
- Début d'apprentissage par interaction (connaissance peut venir de l'interaction).

Environnement inconnu

Si la dynamique du système n'est pas connue, il faut apprendre au cours de l'interaction. On ne peut pas apprendre la stratégie optimale avant d'avoir de l'information sur le système. C'est ce qui définit l'apprentissage par renforcement.

Méthode Naïve : DP Adaptative

Apprendre la dynamique du système par l'interaction (comptage) et appliquer la programmation dynamique.

Méthode Monte Carlo

Apprendre $V^\pi(s)$ par échantillonnage

- Choix aléatoire d'un état $s \in S$
- Suivre la stratégie π et observer le gain R_t
- Faire cela une infinité de fois et moyenner :
  $V^\pi(s) = E^\pi[R_t]$

Méthode Monte Carlo

Apprendre $V^\pi(s)$ par échantillonnage

- Choix aléatoire d'un état $s \in S$
- Suivre la stratégie π et observer le gain R_t
- Faire cela une infinité de fois et moyenner : $V^\pi(s) = E^\pi[R_t]$

Apprendre $Q^\pi(s, a)$ par échantillonnage

- Choix aléatoire d'un état $s \in S$
- Choix aléatoire de l'action $a \in A$ (*exploring starts*)
- Suivre la stratégie π et observer le gain R_t
- Faire cela une infinité de fois et moyenner : $Q^\pi(s, a) = E^\pi[R_t]$
- Améliorer la politique : $\pi(s) = \text{argmax}_{a \in A} Q^\pi(s, a)$

Problèmes

Programmation Dynamique

- Nécessite la connaissance de la dynamique du système
- Mais tiens compte de la strucure :

$$\forall s \in S \quad V^*(s) = \max_{a \in A} E(r(s,a) + \gamma \sum_{s' \in S} \mathcal{T}^a_{ss'} V^*(s'))$$

Monte Carlo

- Aucune connaissance nécessaire
- Mais aucune prise en compte de la structure du MDP : $Q^\pi(s,a) = E^\pi[R_t]$
- Donc, il faut attendre la fin de l'interaction pour améliorer la politique

Différences Temporelles

Principe de la différence temporelle

$$\begin{aligned} V(s_t) &= r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \dots \\ &= r_{t+1} + \gamma V(s_{t+1}) \end{aligned}$$

Mise à jour selon une équation de type Widrow-Hoff :

$$V(s_t) \leftarrow V(s_t) + \alpha(r_{t+1} + \gamma V(s_{t+1}) - V(s_t))$$

- α est le pas d'apprentissage
- $\delta = r_{t+1} + \gamma V(s_{t+1})$ est la cible

SARSA

Idem pour Q

$$Q^\pi(s,a) \leftarrow Q^\pi(s,a) + \alpha(r_{t+1} + \gamma Q^\pi(s',a') - Q^\pi(s,a))$$

`SARSA`

`Initialiser` Q_0

pour $n \leftarrow 0$ **jusqu'à** $N_{tot} - 1$ **faire**

  $s_n \leftarrow$ `ChoixEtat`

  $a_n \leftarrow$ `ChoixAction` $= f(Q^{\pi_t}(s,a))$

  Réaliser action a et observer s', r

  début

    Choisir action $a' = f(Q^{\pi_t}(s',a'))$

    $\delta_n \leftarrow r_n + \gamma Q_n(s'_n, a') - Q_n(s_n, a_n)$

    $Q_{n+1}(s_n, a_n) \leftarrow Q_n(s_n, a_n) + \alpha_n(s_n, a_n)\delta_n$

    $s \leftarrow s'$, $a \leftarrow a'$ **fin**

fin pour

retourner $Q_{N_{tot}}$

Q-Learning

Apprendre π^* en suivant π_t

$$Q^\pi(s,a) \leftarrow Q^\pi(s,a) + \alpha(r_{t+1} + \gamma \max_{a'} Q^\pi(s',a') - Q^\pi(s,a))$$

```
Algorithme Q-learning
```

pour $n \leftarrow 0$ **jusqu'à** $N_{tot} - 1$ **faire**

  $s_n \leftarrow$ `ChoixEtat`

  $a_n \leftarrow$ `ChoixAction`

  $(s'_n, r_n) \leftarrow$ `Simuler`(s_n, a_n)

  % *Mise à jour de* Q_n

  début

    $Q_{n+1} \leftarrow Q_n$

    $\delta_n \leftarrow r_n + \gamma \max_b Q_n(s'_n, b) - Q_n(s_n, a_n)$

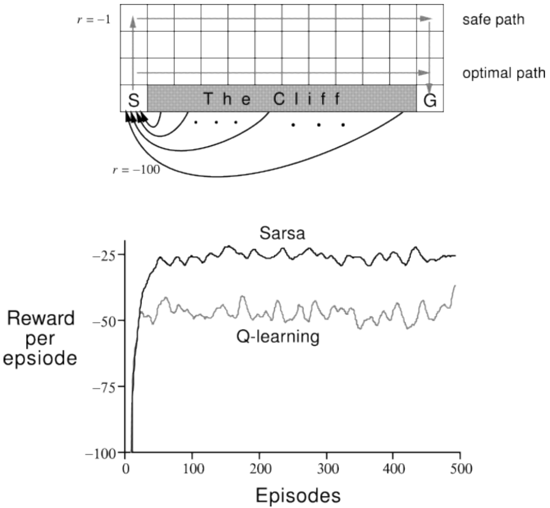
    $Q_{n+1}(s_n, a_n) \leftarrow Q_n(s_n, a_n) + \alpha_n(s_n, a_n)\delta_n$

  fin

fin pour

retourner $Q_{N_{tot}}$

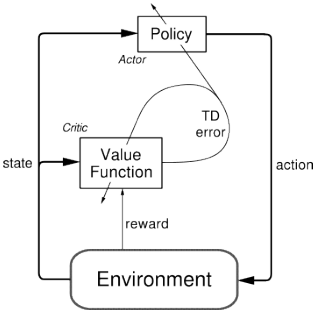
Q-Learning



$r = -1$
safe path
optimal path
S
The Cliff
G
$r = -100$
Sarsa
Q-learning
Reward per epsiode
−25
−50
−75
−100
0
100
200
300
400
500
Episodes

Actor-Critic

Séparation de la représentation de la valeur et de la stratégie. Il s'agit d'une méthode On-Policy.



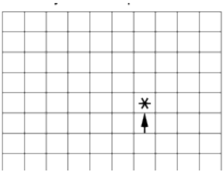
Change la probabilité de la dernière action en fonction de l'erreur TD = $\delta_t = r_{t+1} + \gamma V^{\pi}(s_{t+1}) - V^{\pi}(s_t)$:

$$\pi_t(s,a) \leftarrow \pi_t(s,a) + \beta\delta_t$$

Problèmes avec les méthodes TD(0)

Problème

Si il y a un nombre restreint d'interactions, la propagation de l'information peut ne pas avoir atteint tous les états.

Ex : grid world. 

Solution ?

Se souvenir de toutes les interactions et les reproduire un grand nombre de fois.

Traces d'éligibilité

Dans les méthodes TD, on se base sur $R_t^1 = r_{t+1} + \gamma V_t(s_{t+1})$

On peut aussi écrire :

- $R_t^2 = r_{t+1} + \gamma r_{t+2} + \gamma^2 V_t(s_{t+1})$
- $R_t^n = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots + \gamma^n V_t(s_{t+n})$

Règle de MAJ

$$\Delta V_t(s_t) = \alpha[R_t^n - V_t(s_t)]$$

Forward view I

On peut aussi faire une mise à jour sur n'importe "moyenne" de différents R_t :

- $R_t^{moy} = 1/2R_t^2 + 1/2R_t^4$
- $R_t^{moy} = 1/3R_t^1 + 1/3R_t^2 + 1/3R_t^3$

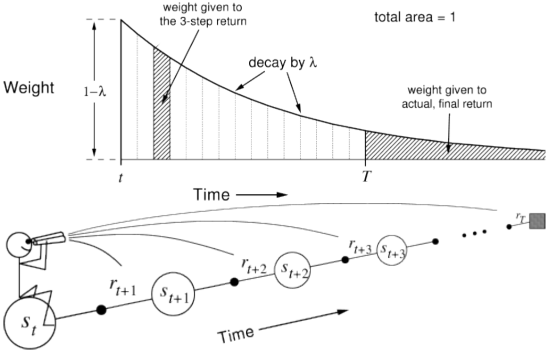
Traces d'éligibilité

$$R_t^\lambda = (1-\lambda)\sum_{n=1}^{\infty} \lambda^{n-1} R_t^n$$

$$\Delta V_t(s_t) = \alpha[R_t^\lambda - V_t(s_t)]$$

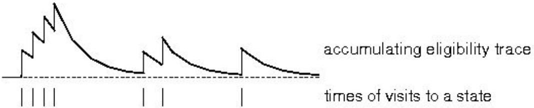
$$0 < \lambda < 1$$

Forward view II



Backward View I

Une variable mémorielle accumulative est associée à chaque état.



$$\forall s, t \; e_t(s) = \begin{cases} \gamma\lambda e_{t-1}(s) & \text{si } s \neq s_t \\ \gamma\lambda e_{t-1}(s) + 1 & \text{si } s = s_t \end{cases}$$

Règle de mise à jour

$$\delta_t = r_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$

$$\forall s \quad \Delta V_t(s) = \alpha \delta_t e_t(s)$$

Backward View II

TD(λ) et Q(λ)

- Tous les états sont mis à jour, le taux d'apprentissage de chacun étant pondéré par la trace d'éligibilité correspondante ;
- si $\lambda = 0$, différence temporelle ;
- si $\lambda = 1$, Monte Carlo

Sarsa(λ)

$$\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$$

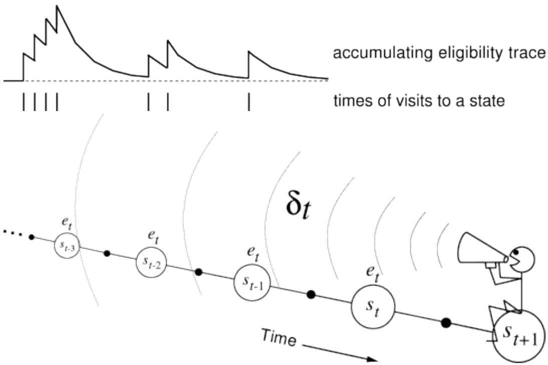
$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t e_t(s, a)$$

Watkin's $Q(\lambda)$

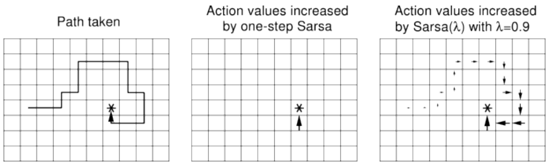
$$\delta_t = r_{t+1} + \gamma \max_{a'} Q_t(s_{t+1}, a') - Q_t(s_t, a_t)$$

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t e_t(s, a)$$

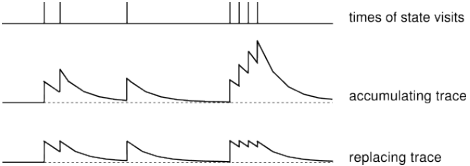
Backward View III

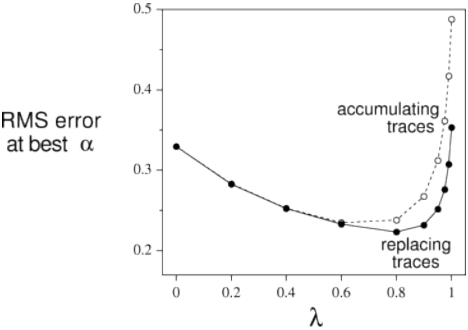


Interprétation

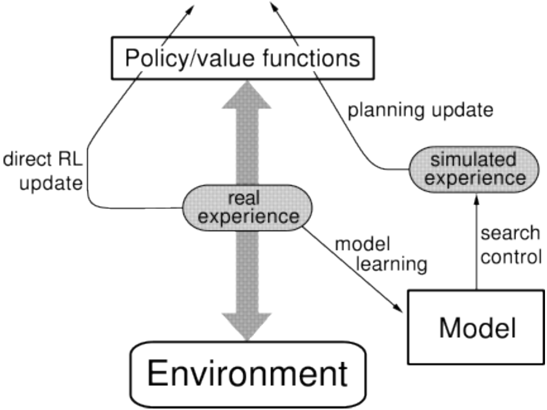


Remplacement des traces



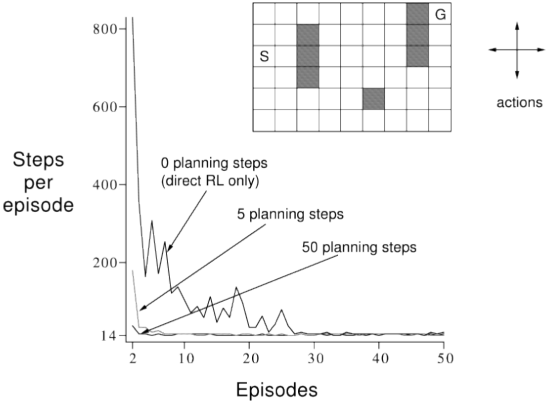


Dyna-Q I



Policy/value functions
planning update
direct RL
update
simulated
experience
real
experience
model
learning
search
control
Model
Environment

Dyna-Q II



G
S
actions
Steps per episode
800
600
400
200
14
0 planning steps (direct RL only)
5 planning steps
50 planning steps
2
10
20
30
40
50
Episodes

Prioritized Sweeping I

Idem Dyna-Q mais choix des paires état-action à mettre à jours ainsi que des probabilités de transition à mettre à jour.

The Curse of Dimensionality (Bellman)

Deux problèmes se posent :

- Comment manipuler de grands espaces d'états/actions ?
- Comment généraliser ce qui a déjà été appris dans des situations semblables lorsqu'on rencontre de nouveaux états ?

Différentes solutions

- *Supervised-learning function approximation* :
  - $s_t \rightarrow v_t = r_{t+1} + \gamma V(s_{t+1})$
  - descente linéaire du gradient,
  - réseaux de neurones, etc...
- *Utilisation de la structure* :
  - MDPs factorisés,
  - MDP hiérarchiques, etc...
- approche multi-agent, etc...

Approximation des fonctions de valeurs

Paramétrisation de la fonction de valeur

On cherche un vecteur de paramètres $\theta \in \mathbb{R}^p$ d'une fonction $\hat{V}_\theta(s)$ ou $\hat{Q}_\theta(s, a)$. La fonction appartient à un espace d'hypothèse $\mathcal{H} = \{\hat{V}_\theta(s) | \theta \in \mathbb{R}^p\}$. Par exemple, la paramétrisation linéaire :

$$\hat{V}_\theta(s) = \sum_{i=0}^{p} \theta_i \phi_i(s) = \theta^T \phi(s)$$

où les $\phi_i(s)$ sont des *fonctions de bases* qui définissent l'espace d'hypothèse.

Question

Comment trouver les paramètres optimaux θ^* ?

Méthodes directes I

Principe général

$$\theta^* = \underset{\theta}{\text{argmin}} \, \| V^\pi(s) - \hat{V}^\pi_\theta(s) \|$$

En pratique : $\theta^* = \text{argmin}_\theta \sum_j \left(v^\pi_j - \hat{V}^\pi_\theta(s_j) \right)^2$ où v^π_j est une observation de $V^\pi(s_j)$

Descente de gradient stochastique

$$\begin{aligned} \theta_i &= \theta_{i-1} - \frac{\alpha_i}{2} \nabla_{\theta_{i-1}} \left(v^\pi_i - \hat{V}_\theta(s_i) \right)^2 \\ &= \theta_{i-1} + \alpha_i \left(\nabla_{\theta_{i-1}} \hat{V}_\theta(s_i) \right) \left(v^\pi_i - \hat{V}_{\theta_{i-1}}(s_i) \right) \end{aligned}$$

Méthodes directes II

Problème

v_i^π n'est pas directement observable (ou bien, pas besoin d'approximation)

Solution : Bootstrapping

On remplace v_i^π par son estimation courante $r_i + \gamma \hat{V}_{\theta_{i-1}}(s_{i+1})$:

$$\theta_i = \theta_{i-1} + \alpha_i \left(\nabla_{\theta_{i-1}} \hat{V}_\theta(s_i) \right) \left(r_i + \gamma \hat{V}_{\theta_{i-1}}(s_{i+1}) - \hat{V}_{\theta_{i-1}}(s_i) \right)$$

Problème

Ca peut diverger si la paramétrisation n'est pas linéaire

Résidus de Bellman I

Principe général

L'estimation $\hat{V}_\theta(s)$ doit vérifier l'équation de Bellman :

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \|\hat{V}^\pi_\theta(s) - B\hat{V}^\pi_\theta(s)\|$$

En pratique : $\theta^* = \operatorname{argmin}_\theta \sum_j \left(\hat{V}^\pi_\theta(s_j) - \hat{T}\hat{V}^\pi_\theta(s_j)\right)^2$ où $\hat{T}\hat{V}^\pi_\theta(s_j) = r_j + \gamma\hat{V}^\pi_\theta(s_{j+1})$

Descente de gradient stochastique

$$\begin{aligned}\theta_i &= \theta_{i-1} - \frac{\alpha_i}{2}\nabla_{\theta_{i-1}}\left(\hat{V}^\pi_\theta(s_i) - \hat{T}\hat{V}^\pi_\theta(s_i)\right)^2 \\ &= \theta_{i-1} + \alpha_i\left(\nabla_{\theta_{i-1}}\left(\hat{V}^\pi_\theta(s_i) - \hat{T}\hat{V}^\pi_\theta(s_i)\right)\right)\left(\hat{T}\hat{V}^\pi_{\theta_{i-1}}(s_i) - \hat{V}^\pi_{\theta_{i-1}}(s_i)\right)\end{aligned}$$

Résidus de Bellman II

Problème

Estimateur biaisé parce que

$$E\left[\left(\hat{V}^{\pi}_{\theta}(s) - \hat{T}\hat{V}^{\pi}_{\theta}(s)\right)^2\right] \neq E\left[\hat{V}^{\pi}_{\theta}(s)\right]^2 - E\left[\hat{T}\hat{V}^{\pi}_{\theta}(s)\right]^2$$

Solution : double échantillonnage

$$\theta_i = \theta_{i-1} + \alpha_i \left(\nabla_{\theta_{i-1}}\left(\hat{V}^{\pi}_{\theta}(s_j) - \gamma \hat{V}^{\pi}_{\theta}(s^1_{i+1})\right)\right)\left(r_i + \gamma \hat{V}^{\pi}_{\theta_{i-1}}(s^2_{i+1}) - \hat{V}^{\pi}_{\theta_{i-1}}(s_i)\right)$$

Point fixe projeté I

Principe général

On définit l'opérateur Π qui projette la fonction $V(s)$ sur l'espace d'hypothèse $\mathcal{H}$ pour obtenir $\hat{V}_\theta(s)$:

$$\Pi V = \underset{\hat{V}_\theta \in \mathcal{H}}{\operatorname{argmin}} \| V - \hat{V}_\theta \|^2$$

On cherche

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \| V(s) - \Pi T \hat{V}_\theta(s) \|^2$$

Point fixe projeté II

Fitted value iteration

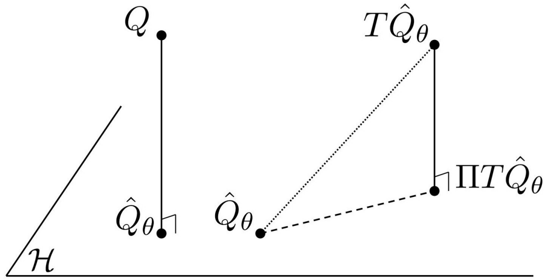
Sous certaines conditions, ΠT est encore une contraction. Il suffit donc de résoudre itérativement :

$$V_{\theta_i} = \Pi T V_{\theta_{i-1}}$$

On doit résoudre ce problème avec des échantillons. On crée donc une base $\{s_j, \hat{T}\hat{V}_{\theta_i}(s_j)\} = \{s_j, r_j + \gamma \hat{V}_{\theta_i}(s_{j+1})\}$ et on résout le problème d'apprentissage supervisé correspondant plusieurs fois itérativement :

$$\hat{V}_{\theta_i}(s_j) = \Pi \hat{T} \hat{V}_{\theta_{i-1}}(s_j)$$

Résumé



- Direct : divergence
- Résidu de Bellman : biaisé
- Projeté : il faut que ΠT soit une contraction

Recherche directe dans l'espace des politiques I

Paramétrisation de la politique

$$\pi^*(s) = \max_a Q(s, a)$$

Si paramétrisation de Q(s,a) : $\pi(s) = \max_a \widehat{Q}_\theta(s, a))$

Problème de discontinuités

Si les actions sont discrètes, la politiques est discontinue et la fonction Q aussi. Un changement dans les paramètres θ donne d'importantes fluctuations de performances et une descente de gradient est impossible.

Recherche directe dans l'espace des politiques II

Solution : politique stochastique

Softmax (Gibbs ou Boltzmann) $\pi_\theta(s,a) = \frac{e^{\widehat{Q}_\theta(s,a)/\tau}}{\sum_{a'} e^{\widehat{Q}_\theta(s,a')/\tau}}$

Montée de gradient

On paramétrise la politique $\pi_\theta(s,a)$. On définit une mesure de performance de la politique η_θ. Ensuite on cherche la politique qui maximise cette mesure :

$$\theta_i = \theta_{i-1} + \alpha_i \nabla_\theta \eta_\theta$$

Problème

La mesure de performance nécessite souvent l'application de la politique au système. Le gradient s'estime alors par méthode de Monte Carlo.

Gestion de l'exploration

Sélection d'une action

- Sélection gloutonne : $a = a^* = \operatorname{argmax}_a Q(s,a)$
- Sélection ϵ-gloutonne : $P(a^*) = 1 - \epsilon$
- Softmax (Gibbs ou Boltzmann) $P(a) = \frac{e^{Q(a)/\tau}}{\sum_{a'} e^{Q(a')/\tau}}$

Initialisation

- Initialisation optimiste : donne une valeur haute à tous les états pour qu'ils soient visités grâce à la sélection d'action.

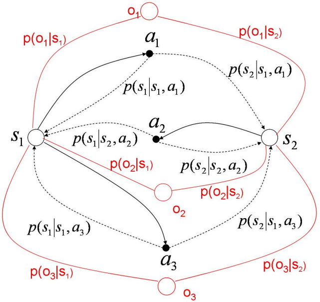
Incertitude et valeur d'information

- Tenir compte de l'incertitude sur Q.
- Calculer la valeur de l'information apportée par l'exploration.

Définition (POMDP)

Un POMDP est constitué de :

- *S l'espace des états ;*
- *A l'espace des actions ;*
- *T l'axe temporel ;*
- *$G \subset S \times A$ le graphe reliant chaque état à différentes actions ;*
- *$(p_t)_{t \in T} = \mathcal{T}^a_{ss'}$ une famille de distributions de probabilité de transitions markoviennes entre actions et états*
- *$(r_t)_{t \in T}$ une famille bornée de fonctions de récompenses sur les transitions d'état*
- *Ω l'espace des observations*
- *O la fonction d'observation sur les états*
- *b_0 la distribution de probabilité initiale sur les états*



interprétation

À chaque instant $t \in T$, l'agent perçoit l'état courant $s \in S$ sous la forme d'une observation $o \in \Omega$ en fonction de la fonction d'observation $O(.)$. Quand il effectue une action a, la dynamique du système l'amène aléatoirement selon $p(.)$ dans un nouvel état s', mais l'agent ne perçoit que $o' = O(a, s')$, et il reçoit une récompense $r(s, a) \in \mathbb{R}$.

Remarques

Si le choix de l'action est déterministe

- Le POMDP devient un Modèle de Markov Caché
- Un POMDP est un modèle de Markov caché *contrôlé*

POMDP = Modèle de l'environnement

- Le POMDP est la modélisation interne du monde faite par l'apprenant
- Si il y a problème de modélisation cela peut être géré par l'intégration de l'insertitude dans le modèle.

Résolution Exacte I

Problématique

- L'agent n'a plus accès à son état et doit choisir ses actions en fonction des observation qui lui sont disponibles.
- Les observations ne constituent pas une statistique suffisante (qui conservent autant d'informations que l'état d'information complet).
- La propriété de Markov est toujours valide sur les états, pas forcément sur les observations.

Solutions

- État d'information complet.
- État de croyance.

Résolution Exacte II

Définition (Etat de croyance ou Belief State))

$b(s) = P(s)$

MàJ belief state

$$
\begin{aligned}
b(s) &= Pr(s) \\
b_o^a(s') &= Pr(s'|b,a,o) \\
&= \frac{P(o|s',a,b))P(s'|b,a)}{Pr(o|b,a)} \\
&= \frac{O(s',a,o)\sum_{s\in S} p(s'|s,a)b(s)}{Pr(o|b,a)} \\
&= \frac{O(s',a,o)\sum_{s\in S} T_{ss'}^a b(s)}{Pr(o|b,a)}
\end{aligned}
$$

Résolution Exacte III

Probabilité de la nouvelle observation

$$
\begin{aligned}
P(o|a,b) &= \sum_{s'} P(o|a,s',b)P(s'|a,b) \\
&= \sum_{s'} O(s',a,o)P(s'|a,b) \\
&= \sum_{s'} O(s',a,o)\sum_{s} P(s'|s,a,b)P(s) \\
&= \sum_{s'} O(s',a,o)\sum_{s} \mathcal{T}^{a}_{ss'}b(s)
\end{aligned}
$$

Résolution Exacte IV

Probabilité entre états de croyance

$$\begin{aligned}\tau(b,a,b') = \tau^a_{ss'} &= P(b'|a,b)\\ &= \sum_o P(b'|o,a,b)P(o|a,b)\\ &= \sum_o P(b'|o,a,b) sum_{s'} O(s',a,o) \sum_s \mathcal{T}^a_{ss'} b(s)\end{aligned}$$

Gain moyen

$$\rho(b|s,s',a) = \rho^a_{ss'} = \sum_s b(s)\mathcal{R}^a_{ss'}$$

Résolution Exacte V

MDP sur Belief States

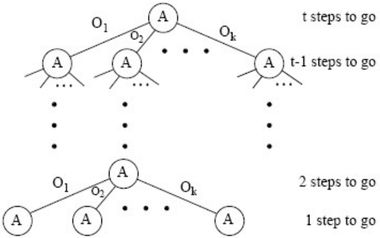
POMDP sur les états $\Rightarrow$ MDP sur les belief states avec

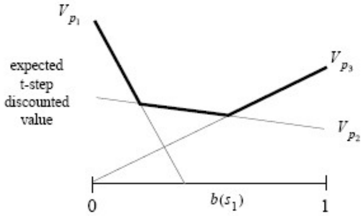
- $\mathcal{T}_{ss'}^{a} \Rightarrow \tau_{ss'}^{a}$
- $\mathcal{R}_{ss'}^{a} \Rightarrow \rho_{ss'}^{a}$

Problème

- Le MDP créé est continu sur b(s) qui est une densité de probabilité !
- Une politique $\pi(b)$ est alors définie sur des région de b(s).
- Algorithme Witness

Résolution Exacte VI





Théorème

La fonction de valeur, à horizon fini, est linéaire par morceau et convexe : $V_t(b) = \max_{\theta \in \Theta_t} b.\theta$

La résolution exacte consiste à générer *intelligemment* Θ_{t+1} à partir de Θ_t. Limité à $|S|$ et $|\Omega|$ d'un ordre de magnitude ($|\Theta_{t+1}| = \mathcal{O}(|A||\Theta_t|^{|\Omega|}$, problème PSPACE-dur à NP-dur)).

Model-Based Methods

Recherche de politiques sous-optimales, les paramètres du modèle (fonctions de transition, d'observation et récompenses) étant connus.

- Approximation de la fonction de valeur
  - Heuristique (par ex. MLS, Most Likely State)
  - Grid Methods (échantillonnage de l'espace de croyance)
- Factorisation ou représentation structuré des états de croyance
  - Utilisation de réseaux bayésiens temporels/dynamiques
- Recherche directe dans l'espace des politiques
  - *Policy Graphs, Finite State Controller...*

Model-Free Methods

Recherche de politiques sous-optimales, les paramètres du modèle (fonctions de transition, d'observation et récompenses) étant inconnus.

- Politiques adpatées
  - ex. `Q-learning` sur les observations plutôt que sur les états
- Apprendre le modèle
  - HMM (Baum-Welch et adaptations)
  - state merging methods
- Méthodes basées sur la mémoire
  - historiques à horizon fini (fixe ou adaptatif -UTREE-)
  - Finite State Controllers
- *Policy-Gradient Methods*

Autres

Certaines approches peuvent s'appliquer à différentes méthodes et potentiellement s'y avérer utile.

- Approche multi-agent
- Approche hiérarchique