

# Projet Foot

## Cours 4

### (Semaine 5)

### 2I013

Nicolas Baskiotis

[nicolas.baskiotis@lip6.fr](mailto:nicolas.baskiotis@lip6.fr)

Université Pierre et Marie Curie (UPMC)  
Laboratoire d'Informatique de Paris 6 (LIP6)

S2 (2014-2015)

# Plan

Résultats de la semaine

Apprentissage Artificiel (Machine Learning)

Les problématiques générales

Premiers modèles : Arbres de décision

# Tournoi 1v1

```
[('Benlog', 'Benlog', 'team_1vs1_Precepteur', 15),  
 ('andrenasturas', 'andrenasturas', 'team1', 12),  
 ('timotheb', 'timotheb', 'team1', 10),  
 ('mariene', 'mariene', 'Poireaux', 5),  
 ('SebXIII', 'SebXIII', 'Pandragon_1v1', 23),  
 ('ArezkiSky', 'Minute_Maid', 'Minute_Maid_Tropical', 3),  
 ('Leynad', 'Maccabi', 'Goooo_1v1', 4),  
 ('Maumiz', 'Maumiz', 'DiegoMaradona_1v1', 23),  
 ('AlexandreCasanova', 'AlexandreCasanova', 'Cherry', 11),  
 ('Leynad', 'Maccabi', '1v1_def', 20)]
```

# Tournoi 2v2

```
[('Benlog', 'Benlog', 'team_2vs2_Temoin', 18),  
 ('Benlog', 'Benlog', 'team_2vs2_Preception', 21),  
 ('andrenasturas', 'andrenasturas', 'team2', 15),  
 ('timotheb', 'timotheb', 'team2', 21),  
 ('orangemango', 'orangemango', 'team1', 9),  
 ('SebXIII', 'SebXIII', 'Tueur_de_fonceur_2v2', 33),  
 ('timotheb', 'timotheb', 'Relegation', 15),  
 ('mariene', 'mariene', 'Patates', 21),  
 ('ArezkiSky', 'Minute_Maid', 'Minute_Maid_Orange', 16),  
 ('SebXIII', 'SebXIII', 'Hyper_Power_2v2', 33),  
 ('Leynad', 'Maccabi', 'Dat_Fonceur_2v2', 7),  
 ('AlexandreCasanova', 'AlexandreCasanova', 'Cramberrie', 14),  
 ('Maumiz', 'Maumiz', 'C.A_BocaJuniors_2v2', 0)]
```

# Tournoi 4v4

```
[('Benlog', 'Benlog', 'team_4vs4_Prescience', 24),  
 ('andrenasturas', 'andrenasturas', 'team4', 24),  
 ('SebXIII', 'SebXIII', 'Unicorn_of_Love_4v4', 13),  
 ('mariene', 'mariene', 'Tomates', 21),  
 ('ArezkiSky', 'Minute_Maid', 'Minute_Maid_Pomme', 10),  
 ('timotheb', 'timotheb', 'Meet_Your_Maker', 18),  
 ('Leynad', 'Maccabi', 'Mauro_Chupame_La_Pija_4v4', 15),  
 ('AlexandreCasanova', 'AlexandreCasanova', 'Lemon', 10),  
 ('mariene', 'mariene', 'Carottes', 19),  
 ('mariene', 'mariene', 'Aubergines', 18),  
 ('Maumiz', 'Maumiz', 'Argentina_4v4', 6),  
 ('Leynad', 'Maccabi', '4v4_maggle', 6)]
```

# Plan

Résultats de la semaine

**Apprentissage Artificiel (Machine Learning)**

Les problématiques générales

Premiers modèles : Arbres de décision

# Quelques définitions

## Scott, 1983

*Learning is the organization of experience.*

## Herbert Simon, 1983

*Learning is any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population.*

## Marvin Minsky, 1985

*Learning is making useful changes in mind.*

## Riszard Michalski, 1986

*Learning is constructing or modifying representations of what is being experienced.*

## L'apprentissage artificiel :

- étudie les algorithmes qui améliore leur performance sur une tache donnée en fonction de leur expérience.
- fondements mathématiques, informatiques et applications concrètes des systèmes qui apprennent, raisonnent et agissent.

# Apprentissage artificielle en pratique ?



A votre avis, c'est utilisé où ?

*Person of interest, série*

# En image

## Détection de visages

(opencv)



# En image

## Détection de visages

(opencv)



## Mais aussi ...

(betafaceapi.com)



Score: 0.42  
X: 398.67  
Y: 29.66  
Width: 26.79  
Height: 26.79  
Angle: -5.45

age : 37 (16%), gender : male, race : white, chin size : average, color background : 4c5042 (15%), color clothes middle : 3295eb (48%), color clothes sides : 38a9f5 (96%), color eyes : ac8066, color hair : fbf2ea (80%), color mustache : a56855 (65%),  
color skin : dbb5a1, eyebrows corners : extra low, eyebrows position : average, eyebrows size : extra thin, eyes corners : low, eyes distance : average, eyes position : average, eyes shape : extra round, glasses rim : no, hair beard : none, hair color type : blond (80%), hair forehead : yes, hair length : none, hair mustache : thick, hair sides : very thin, hair top : short, head shape : average, head width : extra narrow, mouth corners : low, mouth height : extra thin, mouth width : extra small, nose shape : extra straight, nose width : wide, teeth visible : no [collapse]



Score: 0.57  
X: 216.66  
Y: 155.08  
Width: 28.34  
Height: 28.34  
Angle: 0.95

age : 46 (23%), gender : male, race : white, chin size : extra small, color background : 0c0c0d (36%), color beard : 4a2617 (50%), color clothes middle : a22e55 (82%), color clothes sides : a54031 (74%), color eyes : 966a58, color hair : 655348 (77%), color skin : b98f78, eyebrows corners : average, eyebrows position : extra high, eyebrows size : extra thin, eyes corners : average, eyes distance : close, eyes position : extra low, eyes shape : extra thin, glasses rim : no, hair beard : short, hair color type : brown light (77%), hair forehead : no, hair length : short, hair mustache : none, hair sides : thin, hair top : short, head shape : rect, head width : extra wide, mouth corners : average, mouth height : extra thin, mouth width : average, nose shape : average, nose width : extra narrow, teeth visible : no [collapse]

# En image

## Détection d'objets

teradeep.com, Purdue University



## Tracking

[Fragkiadaki et al. 12], Pennsylvania University



# En texte

## Détection de spam

gmail.com

[Supprimer tous les spams maintenant](#) (les messages se trouvant dans le dossier Spam depuis plus de 30 jours sont automatiquement supprimés)

<input type="checkbox"/>	<input type="checkbox"/> Tatianna	Re: Para os homens - Vai lhe interessar muito!	01:50
<input type="checkbox"/>	<input type="checkbox"/> comebuy	Téléphones les plus compétitifs de Comebuy	22:38
<input type="checkbox"/>	<input type="checkbox"/> Francois	100 raisons de jouer sur Majestic	27 janv.
<input type="checkbox"/>	<input type="checkbox"/> Fund Investigation Bureau	TREAT AS URGENT RIGHT AWAY	27 janv.
<input type="checkbox"/>	<input type="checkbox"/> Mrs Elizabeth Johnson	Hello My Beloved One.	27 janv.
<input type="checkbox"/>	<input type="checkbox"/> Evellyn	Re: Amigo, não está satisfeito com o tamanho? Isto pode te ajudar!	27 janv.
<input type="checkbox"/>	<input type="checkbox"/> Amanda, Amanda (2)	Re: Amigo, o que vc faria com 10cm a mais?	26 janv.
<input type="checkbox"/>	<input type="checkbox"/> Groupe Partouche	Et encore un gagnant au Megapot !	26 janv.
<input type="checkbox"/>	<input type="checkbox"/> Carli, Joshua Daniel	N/A	26 janv.
<input type="checkbox"/>	<input type="checkbox"/> RCH Tournoi	Votre Semaine avec 100000 en Tout	26 janv.
<input type="checkbox"/>	<input type="checkbox"/> Jenmy Klamet	Nicolas Baskiottis F-E...E...L-I...N G..._H_O...R_N-Y?-__G-E-T_L_A_I_D_N_O_W!	26 janv.
<input type="checkbox"/>	<input type="checkbox"/> Jean-Pierre	Les meilleurs casinos pour les joueurs français	25 janv.

Principale	Réseaux sociaux	Promotions	
<input type="checkbox"/> CollierPrenom	Annonce	<b>Spécial St Valentin</b> - 3 Jours Seulement - 15% de Réduction !	X
<input type="checkbox"/> SoftLayer.com	Annonce	<b>Get a Secure Cloud</b> - We've secured the public cloud with private servers, private networks, and full private clouds.	X
<input type="checkbox"/> Booking.com		Last-minute deals for Montréal and London. Get them before they're gone!	28/12/2014
<input type="checkbox"/> Voyages-snfc.com		DERNIERE MINUTE NOUVEL AN : profitez des meilleurs prix !	26/12/2014
<input type="checkbox"/> Impossible		Year's End Clearance - Up to 20% off Film and Accessories	26/12/2014
<input type="checkbox"/> Booking.com		Nicolas – you qualify for at least 20% off places to stay	26/12/2014
<input type="checkbox"/> Communauté d'entraide Gr.		Nicolas, des questions sur vos produits ?	25/12/2014

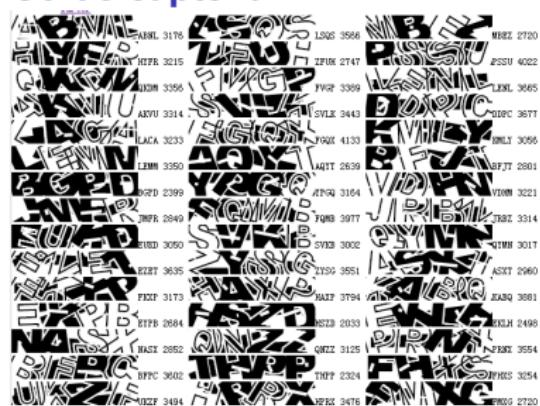
# En texte toujours

## Reconnaissance de chiffres

82944649709295159103  
 23591762822507497832  
 11836103100112730465  
 26471899307102035465

## Ou de captcha

[Yann et al. 08], Newcastle University



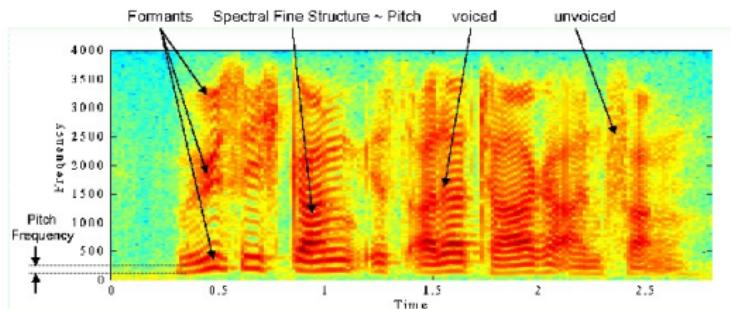
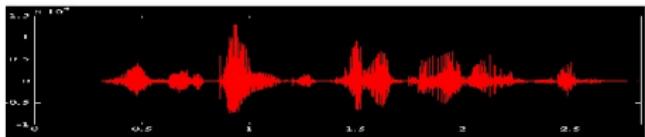
Characters under typical distortions	Recognition rate
5 4 2 2 0	~100%
K K F E B	96+%
P L F C B	100%
D L 1 3 2	98%
W H S R T	~100%
Q D B R	95+%

Et plein d'autres applications (traduction, détection de plagiat, ...)

# Et l'audio ...

## Reconnaissance de la parole

<http://markus-hauenstein.de>



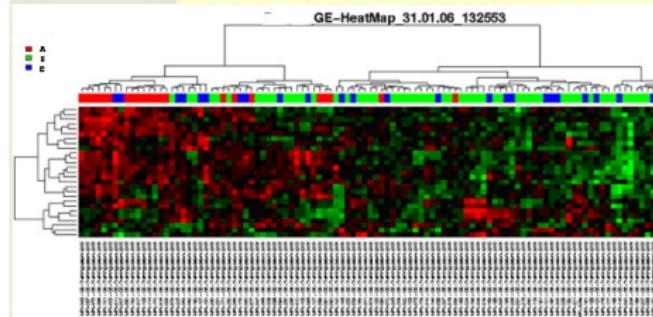
Mais aussi séparation de sources, transcription musicale, ...

# Et dans d'autres sciences

## Biologie IZBI, Leipzig University

### Gene Signal Value Visualization - Gene Expression Heatmap

This form draws the heatmap of Gene Expression signals determined by a selected Experiment Group and a selected Gene Group.



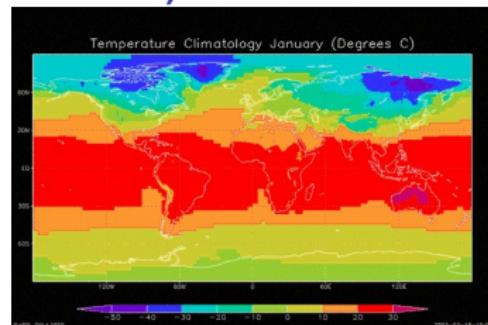
## Économie



## Astronomie



## Climatologie (complémentation données)



# Dans les jeux et la robotique



# Plan

Résultats de la semaine

Apprentissage Artificiel (Machine Learning)

Les problématiques générales

Premiers modèles : Arbres de décision

# L'apprentissage

(selon M. Sahani, UCL)

## En quelques mots

- Trouver des structures, des régularités dans des observations.
- Prédire de nouvelles observations.

## Touche à beaucoup de domaine, interdisciplinarité très forte

- Statistiques : théorie de l'apprentissage, fouille de données, inférence
- Informatique : IA, vision, RI
- Ingénierie : signal, contrôle, robotique
- Science cognitive, psychologie, neuroscience, épistémologie
- Économie : théorie de la décision, théorie des jeux

## L'apprentissage artificiel

- étudie les algorithmes qui améliorent leur performance sur une tâche donnée en fonction de leur expérience.
- fondements mathématiques, informatiques et applications concrètes des systèmes qui apprennent, raisonnent et agissent.

# Quand appliquer l'apprentissage ?

## Lorsque :

- l'expertise humaine est absente
- impossible d'expliquer cette expertise
- les solutions sont dynamiques
- les solutions doivent être adaptées à beaucoup de cas spécifiques
- la taille du problème est trop grand pour que l'humain puisse le résoudre

# Apprentissage supervisé

## Données du problème

- Une représentation  $X$  des objets de l'étude
- Une sortie d'intérêt  $y$  qui peut être numérique, catégorielle, structurée, complexe (label, réponse, étiquette, ...)
- Un ensemble d'exemples, d'échantillons, sous leur représentation  $X$  et avec leur sortie connue  $\{(x_1, y_1), \dots, (x_n, y_n)\}$

## Objectifs

- Apprendre une fonction qui *généralise* les exemples
- Prédire de manière précise la sortie  $y$  pour un nouvel exemple  $x$  non vu
- Comprendre quels facteurs influencent la sortie
- Évaluer la qualité de nos prédictions

# Apprentissage non supervisé

## Données du problème

- Une représentation  $X$  des objets de l'étude
- Un ensemble d'exemples, d'échantillons, sous leur représentation  $X$ ,  $\{x_1, \dots, x_n\}$
- Pas de variable de sortie !

## Objectifs

- Trouver des groupes d'objets “semblables”
  - Organiser les données d'une manière “logique”
  - Trouver les “similarités” des objets
  - Trouver des “représentations” des objets
- ⇒ on ne sait pas bien ce que l'on cherche  
⇒ tout un art !

# Apprentissage par renforcement

Apprentissage continu en fonction du retour d'expérience

## Données du problème

- Un état décrit l'environnement courant
- Un ensemble d'actions sont possibles
- Une politique permet de choisir en fonction de l'état l'action à effectuer
- A l'issue de chaque action, une récompense est observée

## Objectifs

- S'améliorer ! (améliorer la politique de choix de l'action)
- Éviter les situations d'échecs
- Comprendre la dynamique du problème

# L'apprentissage aujourd'hui : Big Data

- Webpages (content, graph)
- Clicks (ad, page, social)
- Users (OpenID, FB Connect)
- e-mails (Hotmail, Y!Mail, Gmail)
- Photos, Movies (Flickr, YouTube, Vimeo ...)
- Cookies / tracking info (see Ghostery)
- Installed apps (Android market etc.)
- Location (Latitude, Loopt, Foursquared)
- User generated content (Wikipedia & co)
- Ads (display, text, DoubleClick, Yahoo)
- Comments (Disqus, Facebook)
- Reviews (Yelp, Y!Local)
- Third party features (e.g. Experian)
- Social connections (LinkedIn, Facebook)
- Purchase decisions (Netflix, Amazon)
- Instant Messages (YIM, Skype, Gtalk)
- Search terms (Google, Bing)
- Timestamp (everything)
- News articles (BBC, NYTimes, Y!News)
- Blog posts (Tumblr, Wordpress)
- Microblogs (Twitter, Jaiku, Meme)



>10B useful webpages

Carnegie Mellon University

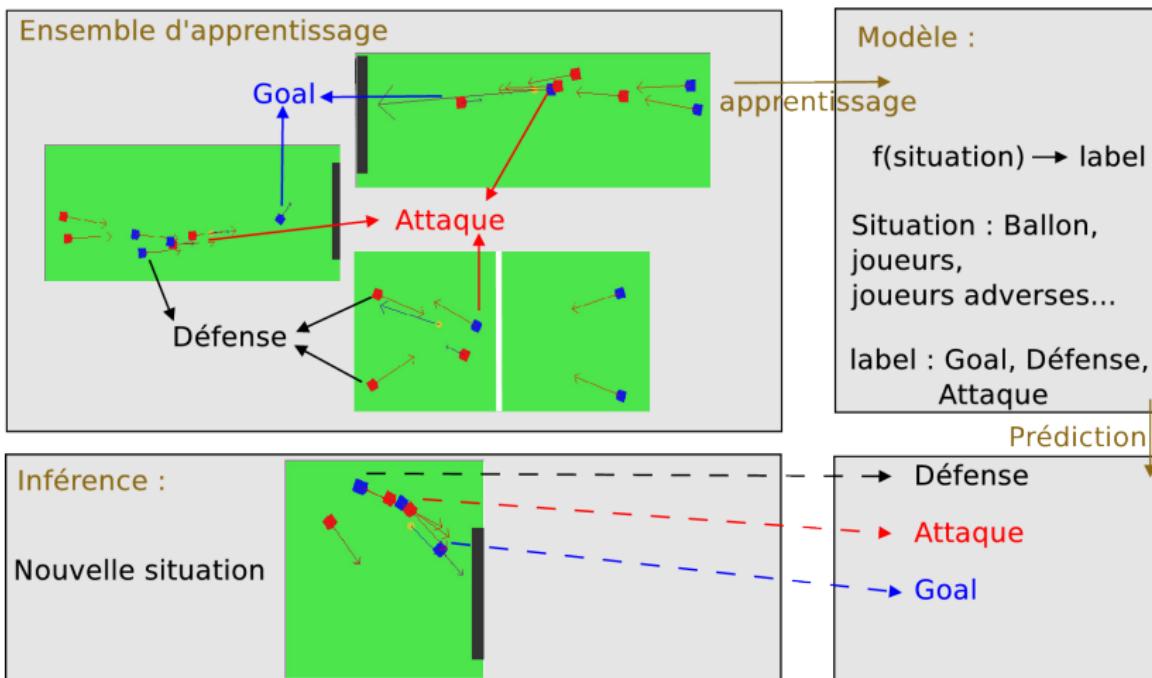
extrait du cours d'A.Smola

## Entreprises concernées :

Yahoo, Google, Amazon, Netflix, Microsoft, Xerox, Critéo, Facebook, Twitter, Flickr, Instagram, Reddit, Valve, Steam, Deezer, Dailymotion, Youtube, STIF, SNCF, AXA, EDF, GDF-Suez, Veolia, les médias, ...

# Et l'utilité dans le cadre du projet ?

## Apprentissage supervisé



Apprentissage par renforcement et algorithmes génétiques plus tard ...

# Plan

Résultats de la semaine

Apprentissage Artificiel (Machine Learning)

Les problématiques générales

**Premiers modèles : Arbres de décision**

# Formalisation de l'apprentissage supervisé

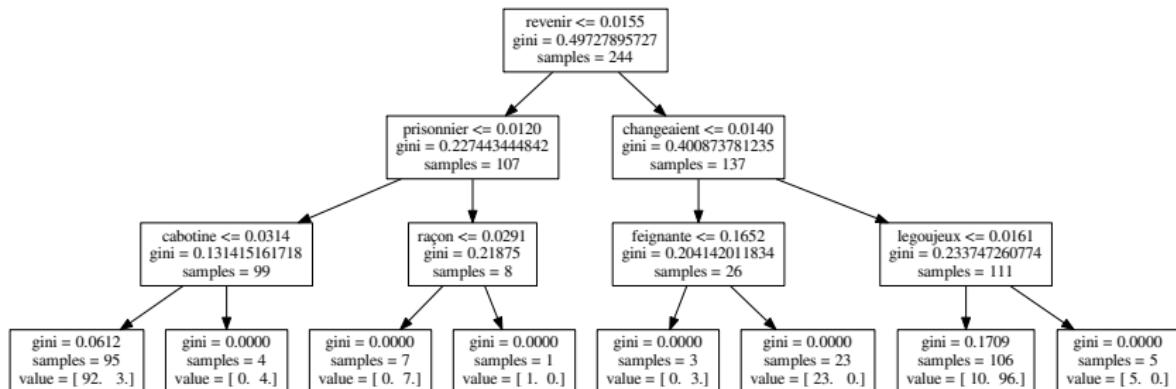
On dispose :

- d'un espace de représentation  $\mathcal{X}$ , usuellement  $\mathbb{R}^n$   
 $n$  est la dimension de l'espace de représentation,  
chaque dimension = un attribut  
⇒ une dimension décrit un élément de la situation
- d'un ensemble d'exemples  $X$  décrit dans cette espace :  
 $x \in X, x = (x_1, x_2, x_3, \dots, x_n)$   
⇒ un exemple = une situation
- d'un ensemble d'étiquettes/labels  $Y$  décrivant les classes  
⇒  $Y$  = ensemble des stratégies possibles
- pour chaque exemple  $x^i$  de  $X$ , son étiquette  $y^i$   
⇒ ensemble d'apprentissage  $E = \{(x^i, y^i)\}$

On veut :

Trouver une fonction  $f : \mathcal{X} \rightarrow Y$  telle que la prédiction sur de futurs exemples soit la plus précise possible.

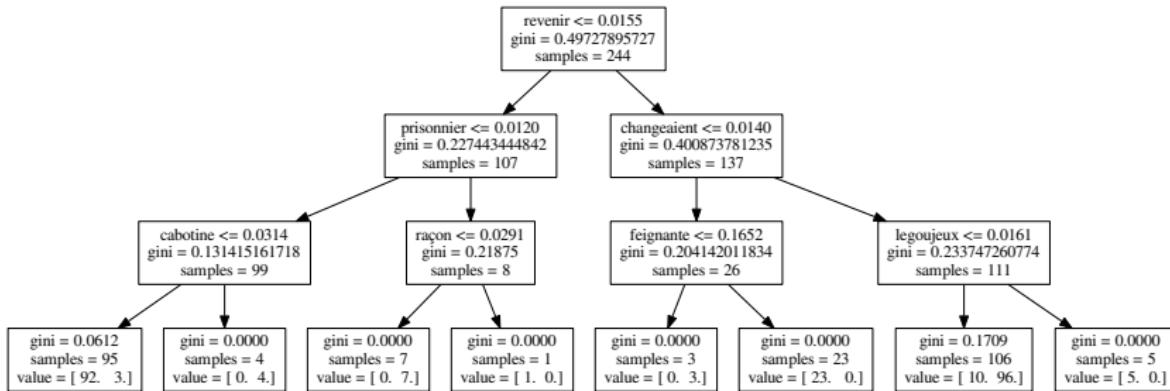
# Arbres de décision



## Principe

- Chaque nœud interne : un test sur une des dimensions de  $\mathcal{X}$
  - Chaque branche : un résultat du test
  - Chaque feuille : un label de  $Y$
- ⇒ classification en parcourant un chemin de la racine à une feuille.

# Arbres de décision

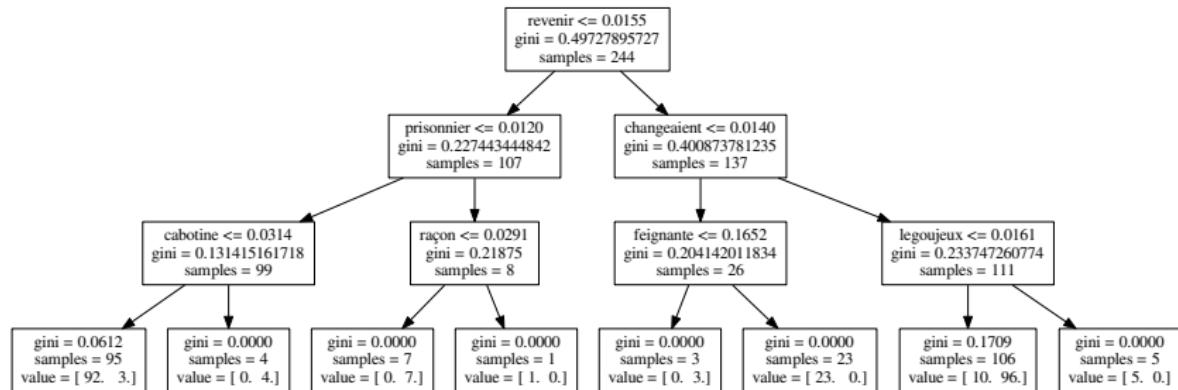


## Exercice

Comment représenter :

- $balle > 1 \wedge adversaire < 0.5$  ?
- $balle > 1 \vee adversaire < 0.5$  ?
- $balle > 1 \wedge adversaire < 1 \vee balle < 1 \wedge equipier > 1$  ?

# Apprentissage d'un arbre de décision

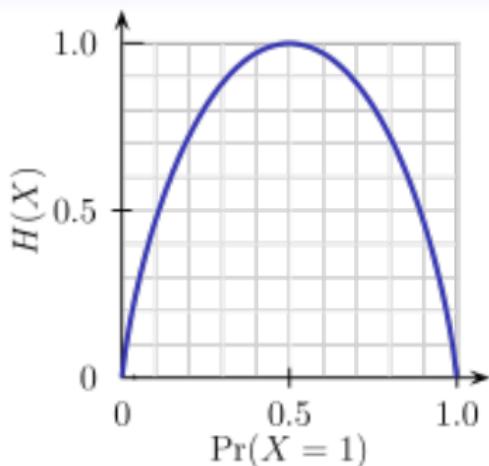


## Algorithme glouton, top-down

Initialisation à la racine, considérer tous les exemples

- Si le nœud n'est pas pur, alors
  - Trouver  $x_i$  le "meilleur" attribut pour ce nœud et le seuil
  - Pour chaque test, créer un fils au nœud courant
  - Affecter les exemples du nœud courant au fils correspondant
- sinon transformer le nœud en feuille.

## Sélectionner le meilleur attribut



### Entropie d'une variable aléatoire

Soit  $X$  une variable aléatoire pouvant prendre  $n$  valeurs :

$$H(X) = - \sum_{i=1}^n P(X = u_i) \log(P(X = i))$$

Plus l'entropie est grande, plus le désordre est grand.  
Entropie nulle  $\rightarrow$  pas d'aléa.

# Sélectionner le meilleur attribut

## Entropie d'un échantillon : cas binaire

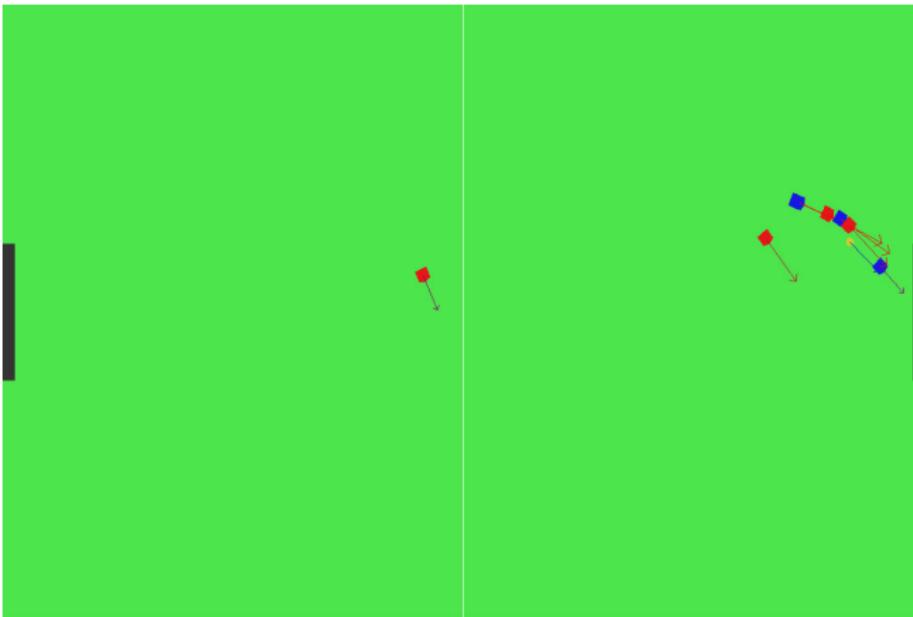
- $X$  un ensemble de données
- $p_+$  la proportion d'exemples positifs
- $p_-$  la proportion d'exemples négatifs
- $H(X) = -p_+ \log(p_+) - p_- \log(p_-)$

## Cas général : entropie conditionnelle

- $H(X|Y = y) = -\sum_{i=1}^n P(X = i|Y = y) \log P(X = i|Y = y)$
  - $H(X|Y) = \sum_{y \in Y} P(Y = y) H(X|Y = y)$
- ⇒ Gain d'information :  $I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

# Description d'une situation

La clé pour un apprentissage réussi ! Que proposez-vous ?



- Attention : un attribut doit pouvoir généraliser !!
- Description : (position joueur, position balle, ...) pas bien → pourquoi ?
- (distance a la balle, au but, distance adversaire) bien ! → pourquoi ?