

Intelligence Artificielle

La créativité algorithmique



L'intelligence artificielle est le mot clé de notre époque, il est omniprésent dans les discours des banquiers, des hommes politiques, comme dans celui des mathématiciens ou des philosophes. Mais si le concept séduit beaucoup d'experts autoproclamés, le discours médiatique semble finalement assez loin de l'emploi réel de ces technologies.

De fait, l'IA est déjà omniprésente dans notre quotidien, mais pour l'instant elle ressemble beaucoup plus à un classeur Excel qu'au robot exterminateur d'homme du film Terminator. Nous allons donc tenter de revenir à des éléments de compréhension concrets des technologies et enjeux qui entourent l'IA.

Pour commencer, si le large champ de recherche qu'est l'intelligence artificielle fascine autant, c'est parcequ'il questionne **l'intelligence** elle-même. De manière extrêmement sommaire, ce qui va caractériser l'intelligence c'est une capacité à **percevoir** des sons, des images, des odeurs, à les **analyser** (c'est-à-dire à les classer, comparer) et à en tirer des conclusions, **agir** en conséquence.

Le concept de l'intelligence artificielle réside dans l'intégration de ces processus cognitifs à des **machines**. Ainsi, sous cette étiquette englobante, on pourra aussi bien trouver les technologies d'analyse d'image, de vidéo ou de la parole que des outils d'analyse financière.

Le développement de ces technologies se fait depuis de nombreuses années déjà, mais l'explosion du nombre de données, les améliorations de leur traitement algorithmique et l'augmentation de la puissance de calcul offrent de nouvelles perspectives qui sont pour le moins impressionnantes. Internet regorge d'anecdotes sur l'intelligence artificielle. Quand une IA ne peint pas un nouveau [Rembrandt](#), elle réalise des [fausses vidéos](#) présidentielles ou détecte des cas inattendus de [cancer](#).

A l'instar du cloud ou de l'internet des objets, l'intelligence artificielle tient donc une large part de l'agenda des plus **grands groupes technologiques**. IBM a tout misé sur « Watson » son système global d'intelligence artificielle, Facebook a créé un pôle dédié, appelé FAIR. Un centre de recherche en France qui est notamment animé par Yann Le Cun, expert internationalement reconnu sur le sujet. Le groupe Alphabet (Google) a une politique de rachat extrêmement volontariste sur la question. Parmi ces « prises » on peut citer notamment Deepmind, Moodstocks ou encore Gradient Ventures. En ce qui concerne Amazon, l'intelligence artificielle représente d'ores et déjà un tiers de leur chiffre d'affaires.

Phénomène intéressant, les **Etats-Unis**, qui étaient en situation d'hyperdomination, dans le domaine il y a encore quelques années sont petit à petit en train de perdre du terrain. La **Chine** notamment cherche à se positionner sur cette niche. Le soutien se fait à la fois par investissement massif, aussi bien par l'Etat que par les grands groupes comme Alibaba, Tencent, Huawei ou Baidu.

La France n'est pas en reste. Le gouvernement a lancé une initiative « France IA », qui a été suivie par un « plan national pour l'intelligence artificielle ». A l'assemblée, c'est le député-médaille de Fields Cédric Villani qui a été chargé d'incarner ce projet. Le but étant de mettre en place un fond abondé par la vente des parts détenues par l'Etat dans des grands groupes. La Région Île-de-France est particulièrement en pointe sur la question et elle a récemment lancé un grand concours pour soutenir les start-ups innovantes du domaine.

Quelques clés techniques de compréhension de la logique de cette tendance permettent de mieux comprendre ces enjeux stratégiques et l'impact.

Le dataset : La perception par la donnée

Lorsqu'on parle « d'intelligence artificielle » on se base sur une conception empiriste de l'intelligence, qui découle d'Aristote. C'est-à-dire une vision dans laquelle l'intelligence n'est pas une faculté « innée » comme le serait l'intelligence platonicienne, mais un processus d'apprentissage qui se développe à travers un tri rationnel d'informations issues de la perception par les sens. Dans la conception aristotelicienne, il y a donc cette idée de vouloir s'approprier une situation en produisant des masses de connaissances descriptives.

L'idée de récolter, trier et conserver des informations pour rendre intelligible une situation, un phénomène et agir en conséquence, n'est donc pas un phénomène nouveau. C'est à la fois un socle de base pour l'interprétation scientifique, mais aussi un moyen pour les organisations de rationaliser leur action. Les Etats par exemple possèdent des « réserves » de données d'une grande qualité et d'une grande diversité (des données macroéconomiques, les données météorologiques, agricoles, les données minières...).

L'ouverture de toutes ces données dans le cadre de la loi CADA, qui va de paire avec le développement d'internet ouvre un immense champ des possibles pour l'intelligence artificielle.

Celle-ci va se nourrir **de jeux de données ou « dataset »** qui vont lui permettre de construire une représentation interne d'une situation et de développer de nouvelles perspectives et de nouvelles possibilités pour y répondre.

Assez logiquement, ces jeux de données sont constitués à partir de données. Mais ce qui semble simpliste au premier abord mérite néanmoins de rentrer dans le détail pour comprendre l'intérêt de la démarche. Les données dont on parle, ce sont des éléments, comme un classeur, un texte, une image qui peuvent provenir de plusieurs types de source.

Le plus simple ce sont les données qui proviennent de **fichiers « plats »**, c'est-à-dire des fichiers comme ceux qu'on utilise tous les jours (.pdf ; .jpeg ; .csv) qui ne dépendent pas des logiciels d'application et pour lequel il n'y a pas de **liens « structurés »** entre les différents enregistrements. L'avantage de ces formats de données c'est qu'ils parlent à tout le monde, mais ils restent compliqués à exploiter à grande échelle.

Ainsi des langages de description comme XML ou J-JSON (pour java script notation), vont permettre de mettre des balises sur les informations contenues dans le document pour catégoriser la donnée. C'est-à-dire mettre les champs en ordre, préciser leur apparence sur une page, les connections logiques entre les différents champs. C'est par ce biais qu'on va obtenir des fichiers avec des **données dites « structurées »**. Ce qui permet un traitement beaucoup plus simple à une échelle agrégée. Utile pour analyser une masse importante d'information. Pour utiliser un exemple connu, les données de Google Maps sont structurées par un balisage au format « GeoJSON ».

Pour extraire du sens dans une masse de données qui peut très vite être indigeste, les organisations vont chercher à structurer, enrichir et agréger leurs données pour en obtenir des outils de travail adaptés à leurs besoins (tableaux de bord etc..). Pour ce faire, elles constituent des **bases de données** qui prennent une forme différente selon les structures.

La première manière de faire, c'est de mettre en place un « **data-lake** », c'est-à-dire d'avoir un espace dans lequel des données de natures très diverses et surtout en quantité très importantes sont conservées dans leur forme initiale pour être utilisées au gré des besoins et par de nombreux utilisateurs différents. Le concept est largement lié à celui de « Big Data » et il nécessite l'emploi de plateformes adaptées comme **Hadoop ou Amazon Web**

Services pour permettre de constituer un dataset pouvant être traité de manière optimale par les algorithmes de l'IA.

L'avantage de cette démarche, c'est qu'elle permet une grande flexibilité, la réflexion sur l'usage qui est fait sur les données ne se faisant qu'a posteriori. Un datalake permet de ne pas partir de zéro dans le cadre de la création de nouvelles activités, de nouvelles missions. Le problème, c'est qu'une base de données plus large est nécessairement plus complexe à maîtriser.

Une autre manière d'organiser ses bases de données c'est de constituer des entrepôts de données (ou datawarehouse). Ici, le traitement des données est cadré en amont pour faciliter la mise en place d'analyses particulières. Le data warehousing là aussi il est nécessaire d'une plateforme de gestion de données pour classer et trier les données de manière optimale. Parmi les plateformes les plus utilisées, on peut notamment citer Datalogix, Experian, Bluekai ou D&B.

Les algorithmes de machine learning, du dataset à l'analyse

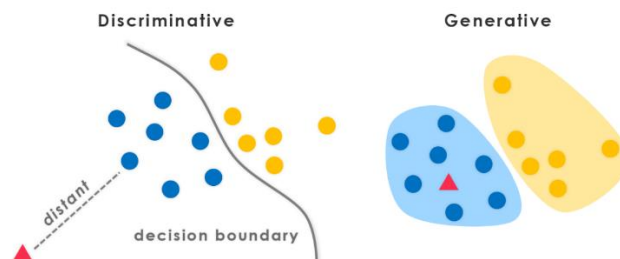
Une fois les données mise en ordre, le véritable intérêt de l'intelligence artificielle, c'est la capacité à adopter un raisonnement logique pour agir de la manière la plus rationnelle possible. A ce titre, Leibniz expliquait dans « de arte combinatoria », vouloir constituer un « alphabet des pensées humaines » dans lequel il se baserait sur des « idées de base » découlant de la déconstruction la plus fondamentale des idées pour ensuite tirer des idées nouvelles par la combinaison de ces structures fondamentales. Une forme d'apprentissage par le calcul.

Plus de trois siècles plus tard, c'est cette logique qui est au cœur des algorithmes qui font l'intelligence artificielle. C'est ce qu'on appelle **l'apprentissage automatique (ou machine learning)**. Pour faire simple, ces algorithmes peuvent être **discriminant ou génératif** et **supervisés ou non**.

Dans une approche **discriminante**, on va chercher à grouper les données qui ont des similarités par catégories. On parle alors de classes. C'est d'ailleurs pour ça qu'on parle d'algorithmes de **classification** (ou classifieurs). Elles sont plutôt utilisées pour les données qualitatives



Dans l'approche **générative**, on analyse par la méthode mathématique de la régression l'interaction de variables les unes par rapport aux autres pour « prédire » où aurait été une donnée qu'on recherche. Elles sont plutôt utilisées pour des données quantitatives.



Dans un apprentissage dit « **supervisé** » le programme s'applique en fonction d'un dataset qui est « étiqueté » (labeled). Concrètement cela veut dire que la signification de la donnée sera associée à celle-ci. A l'inverse, dans un apprentissage « **non supervisé** », le programme n'a pas besoin d'un étiquetage au préalable. Il va déduire de lui même les structures qui sont sous-jacentes à la base de données.

Les algorithmes de classification supervisés

La catégorisation est un processus de base pour des fonctions centrales de l'intelligence humaine. Qu'est ce que le langage sinon une catégorisation d'objets et d'idées ? Qu'est ce que la connaissance sinon la capacité à faire des différenciations ?

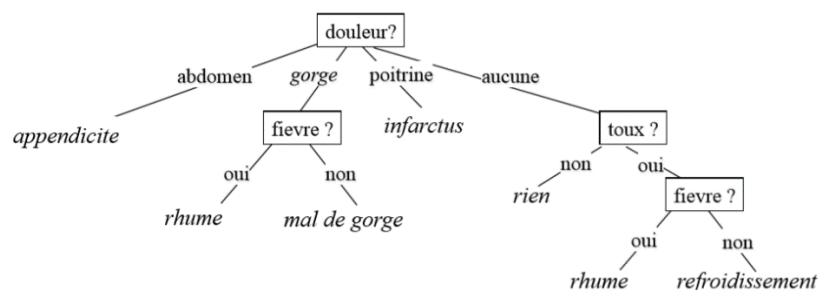
En ce qui concerne l'intelligence artificielle, le processus d'apprentissage se base sur des **classifications « booléennes »** (ou « formelles ») qui vont déterminer des classes de manière binaire (par exemple « oui » ou « non », « malade » ou « pas malade ») et sur des **classifications « bayésienne »** qui vont classer une information en prenant en compte la probabilité d'une hypothèse

Plusieurs types d'algorithmes sont construits sur la base de ces techniques de classification. On compte notamment les arbres de décisions, les réseaux neuronaux.

L'arbre de décision

Dans l'imaginaire collectif, la représentation de différentes catégories d'information sous la forme d'un arbre est relativement classique. Tout le monde a déjà vu un arbre généalogique par exemple.

C'est donc assez naturellement, qu'en informatique, les arbres de décisions comptent parmi les méthodes les plus utilisées en termes de classification. Ils permettent de prendre une décision optimale, lisible et facilement opérationnelle, en prenant en compte plusieurs variables.



Les différents inputs (variables indépendantes) vont permettre de classer une donnée dans des groupes homogènes et la succession de celles-ci permettent d'obtenir des résultats sur une solution ou un comportement à adopter en sortie (variable dépendante).

Si on garde l'exemple du médical, dans le cadre d'une classification « naïve » bayésienne (c'est-à-dire avec des variables indépendantes les unes des autres) on va analyser chaque variable (poids, température, taille) comme une entité autonome, pour calculer un pourcentage de chance que l'individu soit malade ou non.

Cet outil a l'intérêt d'être relativement simple aussi bien dans la compréhension que dans la mise en œuvre. En apprentissage automatique, il est très utilisé notamment pour les chatbots, pour l'analyse de risque, pour l'analyse médicale.

Parmi les arbres de décision les plus utilisés, on peut citer QUEST (qui filtre les variables qui n'ont pas d'importance pour plus d'efficacité), CHAID (qui a une arborescence non-

binaire, utilisé notamment en marketing pour faire des segmentations de marché), ou encore CART (qui intègre des variables de classification et de regression)¹

Si l'usage de ces algorithmes est relativement simple, on obtient un haut niveau de complexité par la mise en place d'un ensemble de plusieurs arbres de décision c'est ce qu'on appelle assez logiquement une forêt d'arbres décisionnels (ou **random forest**). Le fonctionnement de ces algorithmes consiste à proportionner le poids des résultats de chaque arbre pour obtenir un résultat final adéquat. Le « random forest » est utile pour traiter une grande base de données avec peu de prétraitement. Il est utilisé notamment par la X-box Kinect de Microsoft.

Deep learning et réseaux neuronaux

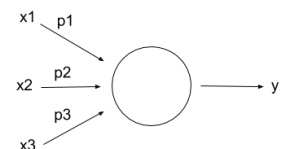
Les exemples les plus fréquemment cités lorsqu'on parle d'intelligence artificielle, sont basés sur le **deep-learning (ou apprentissage profond)** qui fonctionne grâce aux **réseaux de neurones**.

Le vrai pouvoir des réseaux de neurones, c'est qu'ils permettent d'extraire les caractéristiques sous-jacentes d'un jeu de données dit « d'entraînement » (ou *training set*) pour en faire un modèle d'analyse propre à une certaine situation.

Si on prend l'exemple classique de l'utilisation d'un réseau de neurones à convolution pour l'analyse d'image. Il déterminera à partir de données d'images étiquetées « homme » ou « femme » les récurrences présentes dans des visages masculins ou féminins (l'écartement de la bouche, la distance entre les yeux, forme des sourcils, des mâchoires...).

Cette technologie est au cœur des applications concrètes de l'intelligence artificielle et bien loin des grands discours, nous l'utilisons déjà tous les jours. Parmi les cas d'usages les plus courants, on peut citer **Facebook**, qui utilise cette technologie pour la reconnaissance faciale des photos présentes sur son réseau. On peut également citer, **Google** qui l'utilise pour rendre plus précis son moteur de recherche « image ».

Le principe des réseaux neuronaux est d'ailleurs assez ancien, le **perceptron**, qui en est à l'origine date des années 50 ! Concrètement, un neurone artificiel va être une fonction qui va traiter des inputs avec des pondérations différentes (on parle de **poids synaptique**) pour donner un certain résultat en sortie. Le concept est lié à la notion de gradient en mathématiques.



Un « réseau de neurone » va être une combinaison très structurée de ces neurones artificiels dans lequel les informations vont passer d'un neurone à un autre. Comme ça un ensemble d'analyses plutôt simple peuvent amener à des résultats d'une incroyable complexité. Ce traitement de la donnée s'appelle une « **descente de gradient** ».

Initialement, dans les réseaux neuronaux les plus simples, le passage des informations se faisait simplement dans un sens. Les données d'origine subissaient un traitement par différentes couches de neurones pour donner un certain résultat en sortie.

Pour plus de précision, les algorithmes qui sont le plus souvent utilisés lorsqu'on parle de deep-learning vont partir de ce résultat de sortie et revenir réajuster les poids des différents neurones pour que le « coût » c'est-à-dire la différence entre l'output obtenu et l'output souhaité, soit le plus faible possible. C'est ce qu'on appelle la **retropropagation du gradient (ou backpropagation)**. Les avancées en la matière sont notamment dues aux travaux du français Yann Le Cun. Elles ont largement contribué au regain d'intérêt récent

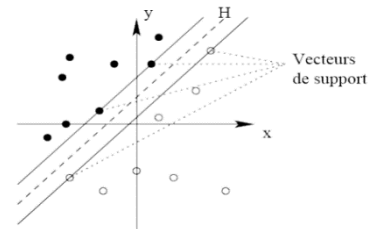
¹ C5.0 et M5 sont également des algorithmes notables en la matière

pour l'IA puisqu'elles ont permis aux réseaux neuronaux de créer des modèles d'une précision extrême.

Le problème du deep-learning, vient du fait qu'il peut vite entrainer un phénomène de « surapprentissage », c'est-à-dire un apprentissage plus précis que de nécessaire au prix d'un besoin important en puissance de calcul. Des besoins importants émergent donc en algorithmique sur la logique d'optimisation et sur la partie hardware vers des processeurs spécifiquement conçu pour le deep-learning.

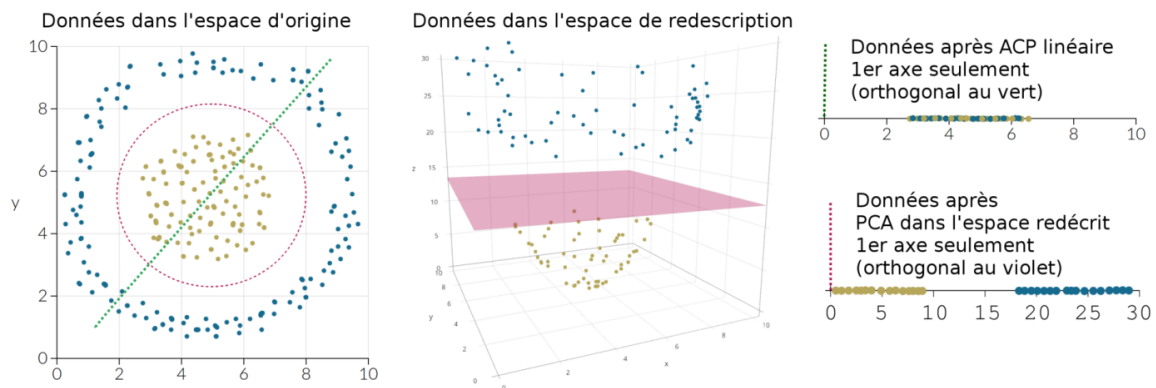
Les machines à vecteur de support (SVM)

Ainsi les machines à vecteurs de support ou « supports à vaste marge » sont de plus en plus utilisés dans tous les segments d'activités concernés par l'IA (image, vidéo, finance).



Ils permettent d'éviter le surapprentissage en se basant sur des « vecteurs de support », c'est-à-dire les éléments à l'extrémité de chaque catégorie. On parle donc également de « support à vaste marge » puisque c'est le fait d'avoir une marge (ou hyperplan) le plus large possible qui permet l'efficacité de ce modèle.

L'autre intérêt des algorithmes SVM, c'est qu'ils permettent de séparer des données de manière « non linéaire ». Pour se faire, on utilise ce qui s'appelle l'astuce du noyau (ou Kernel Trick) c'est-à-dire que les données sont projetées dans un espace où elles deviennent séparables.

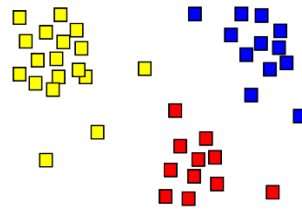


LES ALGORITHMES DE CLASSIFICATION NON-SUPERVISEE : LE CLUSTERING

Le partitionnement de données (ou *clustering* en anglais) consiste à regrouper un ensemble de données qui ont des similarités dans des paquets (ou clusters) pour des données qui n'ont pas été étiquetées au préalable.

En fonction du type d'algorithme utilisé, la notion de ce qu'est un cluster et la manière dont ces groupes vont se constituer va être différente. Il peut s'agir d'une distribution statistique particulière, d'un intervalle particulier ou encore d'une densité particulière dans la répartition des données.

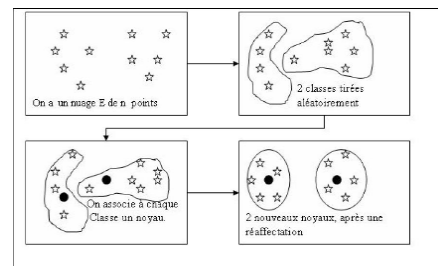
Pour obtenir un bon algorithme de clustering, qui permettra de faire des paquets de données les plus homogènes possible et qui amenera des différenciations les plus nettes entre les différents paquets, on va donc jouer sur les mesures et les classes de distance (comme les quartiles, écarts-types ou variance).



Les algorithmes de clustering vont avoir de nombreux usages, ils vont permettre de segmenter une base de données et d'en extraire des groupes et sous groupes.

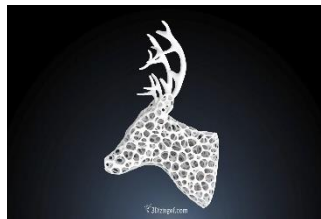
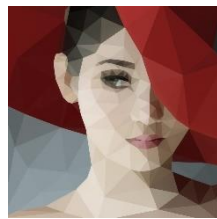
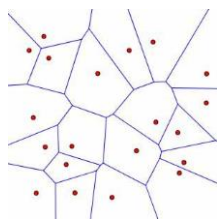
Parmi les exemples d'usage les plus parlant, on peut citer la reconnaissance d'image. Dans l'imagerie satellitaire par exemple, la segmentation par le clustering permet de traiter un nombre important d'informations en se focalisant sur un aspect précis des images étudiées (on se concentre sur les forêts ou les routes par exemple et le traitement ne se fait que sur ce point).

Parmi les méthodes qui sont employés on peut citer les **méthodes basées centroïdes** comme *la partitionnement en k-moyennes* (ou *k-means* en anglais) dans lequel le cluster est constitué d'un groupe de données dont la distance par rapport à la moyenne des autres points du cluster est faible.



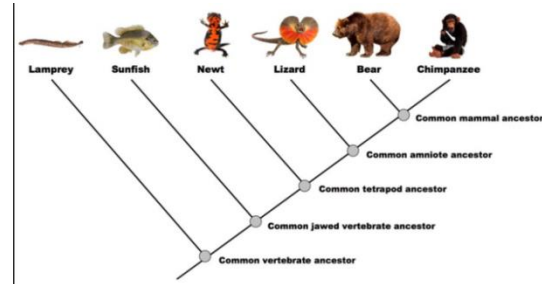
Cette logique est employée notamment dans l'algorithme Colibri-RankBrain utilisé par google pour mieux gérer les requêtes qui n'avait jamais été rentrées auparavant dans le moteur de recherche.

C'est ce type de logique qu'on applique dans le cas des **cellules voronoi** qui sont utilisées aussi bien dans les logiciels de retouche d'image, que dans le design, le diagnostic du cancer ou les atterrisseurs dans le spatial.



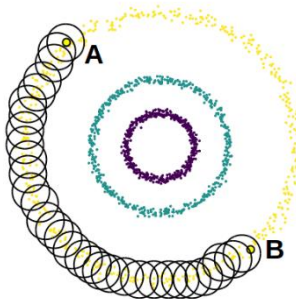
Les méthodes de **regroupement hiérarchique** sont une autre méthode de partitionnement de données.

Les données peuvent alors être classées sous la forme d'un dendrogramme, c'est-à-dire arbre binaire de clustering qui prend la forme du schéma ci-joint.



L'intérêt des algorithmes de regroupement hiérarchique vient notamment du fait qu'ils peuvent intégrer une large base de donnée et qu'il est possible de se contenter d'un nombre de cluster donné (puisque'il s'agit de « couper » le dendrogramme). Comme le montre l'exemple, il est très utile en biologie par exemple.

Par contre il n'est pas le plus simple pour avoir un partitionnement optimal des données puisque'il est sensible aux données atypiques (en dehors de toute catégorie).



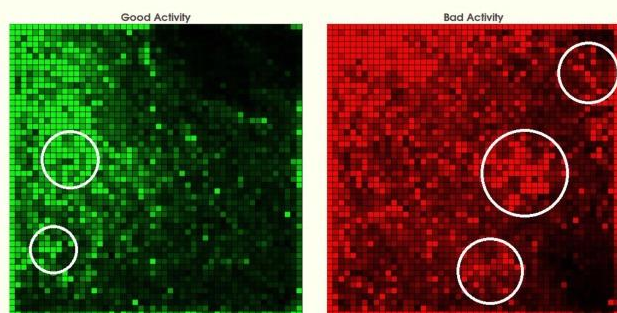
La **méthode de clustering par densité** est également utilisée de manière plutôt classique pour le clustering. Comme son nom l'indique, ces algorithmes (tel que DBSCAN ou OPTICS) permettent de faire des regroupements de données en fonction de leur concentration spatiale dans certaines zones, les clusters sont donc séparés par des zones dites de « bruit » qui ont une moindre densité.

Il existe des modèles de **mélange gaussien**, tel que l'algorithme de maximisation de l'espérance (EM) qui vont combiner plusieurs fonctions de densité. Ils sont très utiles dans la reconstruction d'une base de donnée incomplète.

Les **méthodes connexionnistes** (réseaux neuronaux) sont parfois employés pour effectuer des tâches de clustering (sans supervision), il s'agit des cartes dites « auto adaptatives », également appelés réseaux de Kohonen.

Contrairement aux réseaux de neurones que nous avons précédemment décrit. Ce type d'algorithme fonctionne par « apprentissage compétitif », c'est-à-dire que chaque neurone y est en « concurrence » avec les autres et le système détermine la solution la plus adaptée au traitement de l'input. D'où le surnom d'algorithme « winner-take-all ». Ces neurones sont ensuite ajustés pour correspondre avec encore plus de précision avec l'input.

Après traitement par l'algorithme, les données sont représentées sous la forme d'une « carte » (le plus souvent à deux dimensions) qui permet de voir des similarités entre les données qui ne sont pas visible avant traitement. L'exemple ci-contre montre par exemple l'utilisation d'un réseau de Kohonen pour la détection de malwares.



Prédire avec les algorithmes de regression supervisés.

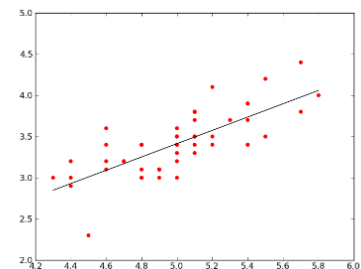
« Quand il croit s'exprimer de façon spontanée, faire œuvre originale, l'artiste réplique à d'autres créateurs passés ou présents, actuels ou virtuels. Qu'on le sache ou qu'on l'ignore, on ne chemine jamais seul sur le chantier de la création » Levi-Strauss dans « la voie des masques »

La créativité, entendue comme la capacité à réaliser une production originale a longtemps été considérée comme une preuve ultime d'intelligence. Au point d'en faire une caractéristique fondamentale de l'être humain. De ce fait, on l'appréhende de manière presque mystique, comme un phénomène qui serait spontané. Pourtant, comme le rappelle Levi-Strauss, la création s'intègre toujours dans un contexte, un cadre. Elle n'est pas un hasard pur. Et par voie de conséquence, les réflexions sur la possibilité de la rendre mathématisable ne sont pas forcément nouvelles.

De fait, en mathématique, lorsqu'on va créer une nouvelle donnée à partir de données existantes au préalable on va appeler ça une prédiction. On l'obtient en passant par une regression.

Régression linéaire

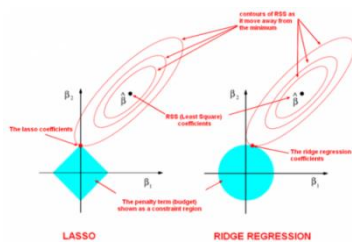
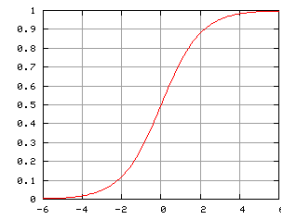
Dans le cas d'une **régression linéaire**, on va vouloir déterminer une droite de régression qui va passer au plus près des points d'un ensemble. Parmi les exemples simples de modèles de regression linéaire, on peut citer **l'ajustement affine** (le fameux $ax+b$) ou encore **la méthode des moindres carrés** (qui calcule la distance de chaque point par rapport à une droite pour avoir un total de l'ensemble de ses distances le plus faible possible)



Mais l'algorithmique de l'intelligence artificielle fait appel à des regressions statistiques qui sont généralement plus complexes.

On peut citer notamment la méthode du **maximum de vraisemblance** ou encore **l'inférence bayésienne**. Qui permettent de faire des régressions non-linéaires.

On peut également citer le **modèle logit**, également très utilisé, c'est une régression qui est faite sur des données qui connaissent une distribution dite « logistique », c'est-à-dire des données qui, comme sur le modèle ci-joint, augmentent rapidement avant de se stabiliser (utilisé notamment pour modéliser l'innovation en économie ou pour les phénomènes de combustion en chimie).



Pour les régressions les plus complexes, il existe des processus dit de « régularisation » qui permettent de simplifier les algorithmes pour optimiser les calculs. On peut citer à ce titre Le LASSO (Least Absolute Shrinkage and Selection Operator) ou la régularisation de Tikhonov (ou ridge regression).

L'utilisation des regression Réseaux de neurones à fonctions radiales (RBF)

L'usage de la logique probabiliste et des régressions mathématiques pour créer de nouvelles données permet de décupler les possibilités des réseaux neuronaux dans la mesure où elle permet de créer des systèmes véritablement prescriptifs. Capables de tirer des conclusions créatives d'un jeu de données.

Les réseaux de neurones à fonctions radiales sont des réseaux de neurones dont les neurones de base vont permettre de créer des nouvelles données après avoir calculé la probabilité de leur justesse. Ce type de réseau de neurones est de plus en plus utilisés parcequ'ils permettent plus de choses que des réseaux neuronaux basés sur/ des classifieurs booléens.

Concrètement, lorsqu'une intelligence artificielle est utilisée pour de la photographie, les algorithmes RBF sont souvent utilisés pour générer les pixels qui vont manquer à cause d'une mauvaise compression, d'un problème de netteté ou de manque de lumière.

Encore une fois, l'exemple de la photographie est un exemple sexy, mais cette logique s'applique à de très nombreux domaines (le traitement du signal, la reconnaissance de la parole, le diagnostic d'erreurs etc...).

La chaîne de markov

Les chaînes de Markov sont également des modèles mathématiques utiles pour générer de nouvelles données sur la base d'une situation existante. C'est un système mathématique qui va passer d'un **état** à un autre de manière aléatoire. Si on prend l'exemple de la musique générée par intelligence artificielle (...) l'état est un accord (représenté sous la forme d'une liste d'intervalles). On va par exemple avoir les accords "Mi mineur" ou "La majeur" comme état, l'ensemble de ces accords forme une liste des états possibles qui s'appelle un "**espace d'état**". La chaîne de Markov va déterminer la probabilité d'une donnée de transiter d'un état vers l'autre

Si on garde l'exemple d'une chaîne simple avec deux états ("La majeur" et "Mi mineur"), 4 transitions seront possibles. En effet une chaîne pourra transiter avec elle même. Si on est dans un état "La majeur" on peut transiter vers "Mi mineur" ou vers "La majeur". Si on est dans un état "Mi mineur" on peut transiter vers un état "La majeur" ou "Mi mineur".

Une matrice de transition permet de représenter la probabilité de chaque transition.

Chaque fois qu'on ajoute un état dans la matrice, on l'ajoute à la fois une ligne ET une colonne. Ce qui amène à des matrices qui sont rapidement très grosses.

Nous avons pris l'exemple de la musique, mais les chaînes de Markov et les algorithmes qui en découlent servent également pour la météo (on regarde la probabilité d'un état pluvieux ou ensoleillé pour déterminer des prévisions) ou pour les moteurs de recherche. ("page-rank" de Google va mesurer avec quelle probabilité un internaute se retrouve sur un site donné après avoir suivi de nombreux liens de manière aléatoire et en tirer un classement de pertinence).

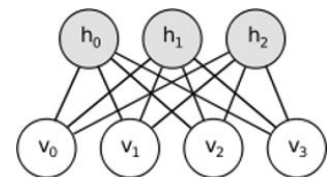
Les algorithmes génératifs non supervisés

Dans un grand nombre de cas il est difficile d'obtenir des modèles pertinents sans avoir un jeu de données suffisamment propre pour le faire. Pour ce faire certains algorithmes permettent de déterminer des solutions à partir de données complexes (brutes, non-structurées ou avec du bruit). Comme les données des satellites d'observations de la Terre, de la génomique (et plus largement l'analyse d'image et vidéo).

Boltzmann machine et Restricted Boltzmann machine

La machine restreinte de Boltzmann notamment est le premier programme de deep learning à fonctionner sur une base de données non supervisée.

Ce réseau de neurones fonctionne avec une « couche visible » et une couche invisible. Dans la première couche, toutes les données ne sont pas forcément étiquetées. Les informations sont ensuite traitées par une couche « invisible », qui va permettre d'avoir un étiquetage correct.



Cet outil est particulièrement efficace dans la préparation d'un bon dataset, en complément d'autres outils algorithmiques.

Les autres outils de compression linéaires non supervisés

Depuis d'autres outils de compression linéaire ont été développés pour extraire des caractéristiques principales et se recentrer sur les variables les plus pertinentes notamment les cartes de Kohonen, l'analyse en composantes principales (PCA) ou la régression des moindres carrés partiels.

CONCLUSION

Revenir aux fondements algorithmiques de l'IA est important pour comprendre la logique du machine learning mais les entreprises et start-ups qui développent de nouvelles applications ne « redécouvrent pas la roue » à chaque développement. Elles peuvent largement prendre appui sur des bibliothèques de fonctions ou de datasets, tel que Scikit-Learn, TensorFlow, Apache MXN ou Caffe2 ils font largement appel à des travaux déjà réalisés, notamment à travers les API (Application programming interface). Ainsi, l'ouverture des API par les grands groupes (Baidu, Microsoft, Facebook, Google) a largement contribué au développement de ces technologies pour de nombreux usages.

Elle permet également de sortir des visions « éditoriales » de l'IA qui est tour à tour montrée comme un futur utopique ou dystopique. Mais si la domination du monde par une supra-intelligence virtuelle relève de la science fiction, l'emploi extensif des technologies de l'intelligence artificielle dans les principales activités est bel et bien une réalité, d'ores et déjà présente sous la forme de tableaux et de formules mathématiques.