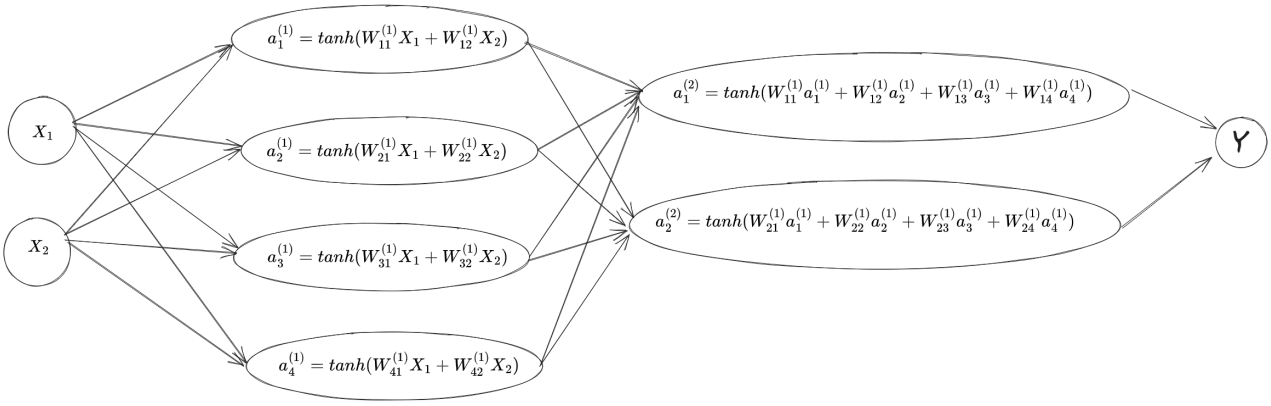
# Partie 1

1. Les étapes de la réalisation d'un projet de Machine Leanrning consistent à comprendre le problème de l'entreprise, à convertir les données dans un format utilisable et à traiter les variables, en soustrayant les variables pertinentes de l'ensemble de données, ce qui équivaut à trouver les variables qui ont une forte relation, à saisir les tendances et avec l'objectif, à effectuer le nettoyage des données pour une analyse primordiale et cohérente, Créer des caractéristiques révélatrices pour améliorer les performances du modèle Diviser ensuite vos données en 3 parties Entraîner la validation et ensuite tester les données, Rechercher le meilleur modèle qui peut résoudre le problème des données, Entraîner le modèle et l'évaluer en utilisant des métriques atteignant le seuil pour bénéficier d'une meilleure décision et d'une meilleure compréhension et c'est avec l'optimisation des données et l'interprétation du résultat et de sa valeur, si tout est fait, nous pouvons déployer garder à l'esprit que le modèle doit être en production, en assurant la maintenance et la fiabilité.
2. Pour préparer les données à partir d'un ensemble de données donné, il est essentiel de comprendre d'abord la nature des variables grâce à l'analyse univariée (histogrammes, diagrammes à barres, mesures de la tendance centrale). Ensuite, l'analyse bivariée explore les relations à l'aide de matrices de corrélation, de diagrammes en boîte et de diagrammes de dispersion. L'analyse des séries chronologiques se concentre sur les tendances et les autocorrélations, tandis que l'analyse géospatiale utilise des cartes thématiques, la corrélation spatiale et le regroupement. L'analyse

transactionnelle étudie les modèles d'achat, les montants des transactions et la segmentation de la clientèle. Enfin, l'ingénierie des caractéristiques crée de nouvelles variables.
Ces étapes, justifiées par leur contribution à la qualité des données, améliorent l'analyse pour des résultats fiables.

3. La conception de l'architecture d'un réseau neuronal dépend fortement du type de données et de la nature de la tâche à accomplir. Pour l'adapter à votre ensemble de données spécifique, vous devrez ajuster le nombre de neurones dans chaque couche, en fonction de la complexité de vos données et de la tâche à accomplir. Par exemple, si vous avez des données tabulaires, vous pouvez avoir une couche d'entrée avec un neurone par variable, et la couche de sortie avec le nombre de classes que vous voulez prédire.




4. Les fonctions d’activation sont utilisées pour reproduire le potentiel d’activation que l’on retrouve dans le domaine de la biologie du cerveau humain. Elles permettent le passage d’information ou non de l’information si le seuil de stimulation est atteint. Les fonctions d’activation les plus populaires sont ReLU, Sigmoid, TanH, Softmax, et Leaky ReLU. Le choix de la fonction d’activation dépend de la tâche à accomplir et de la nature des données. Par exemple, ReLU est souvent utilisé pour les réseaux de neurones convolutifs, tandis que Sigmoid est utilisé pour la classification binaire

Pour cette question je mets en details dans un jupyter notebook leur implementation ainsi que leurs logique mathématique : Deep-Learning-Basics/1-Basics/Activation Function.ipynb at main · IbLahlou/Deep-Learning-Basics (github.com)

5. La généralisation d'un modèle représente sa capacité, une fois entraîné, à effectuer des prédictions sur des données qu'il n'a jamais vues. Pour assurer ce point, il est important de séparer les données en deux ensembles: un ensemble d'entraînement et un ensemble de test. L'ensemble d'entraînement est utilisé pour entraîner le modèle, tandis que l'ensemble de test est utilisé pour évaluer les performances du modèle. Il est également important de s'assurer que le modèle n'est pas sur-ajusté (overfitting) ou sous-ajusté (underfitting) aux données d'entraînement
6. L'overfitting et l'underfitting sont des problèmes courants en apprentissage automatique. L'overfitting se produit lorsque le modèle est trop complexe et s'adapte trop bien aux données d'entraînement, mais ne généralise pas bien aux nouvelles données. L'underfitting se produit lorsque le modèle est trop simple et ne s'adapte pas suffisamment aux données d'entraînement]
7. Le Bias-variance tradeoff est un concept important en apprentissage automatique. Il décrit la relation entre la complexité d'un modèle, l'exactitude de ses prédictions et sa capacité à faire des prédictions sur des données jamais vues auparavant. Un modèle avec une faible complexité aura une forte tendance à sous-ajuster les données, tandis qu'un modèle avec une forte complexité aura une forte tendance à sur-ajuster les données
8. Le k-fold cross validation est une technique de validation croisée qui permet d'évaluer les performances d'un modèle en divisant les données en k groupes (ou "folds") de taille égale. Le modèle est entraîné sur k-1 groupes et testé sur le groupe restant. Cette

procédure est répétée k fois, chaque groupe étant utilisé une fois comme ensemble de test. Le score final est la moyenne des scores obtenus pour chaque groupe

9. Le k-fold cross validation peut résoudre le problème d'overfitting en fournissant une estimation plus précise de la performance du modèle sur des données inconnues. Il peut également aider à assurer un Bias-variance tradeoff optimal en permettant de comparer les performances de différents modèles et de sélectionner le plus approprié pour un problème spécifique
10. Pour améliorer la qualité de chaque modèle, il est important de suivre les pratiques suivantes:

- Collecter des données de haute qualité et en quantité suffisante.
- Nettoyer et normaliser les données.
- Sélectionner les caractéristiques les plus pertinentes pour la tâche à accomplir.
- Utiliser des algorithmes d'apprentissage automatique appropriés pour la tâche à accomplir.