UNIVERSITÉ ABDELMALEK ESSAÂDI

Le modèle mathématique du RNA

Présentée par Soukaina El Messaoui

Plan :

Définition

Les Réseaux de Neurones Artificiels (RNA) sont des modèles inspirés du cerveau humain qui apprennent à partir de données. Leur fonctionnement repose sur des concepts mathématiques comme l'algèbre linéaire, l'optimisation, et les probabilités.

Structure du Réseaux de Neurones

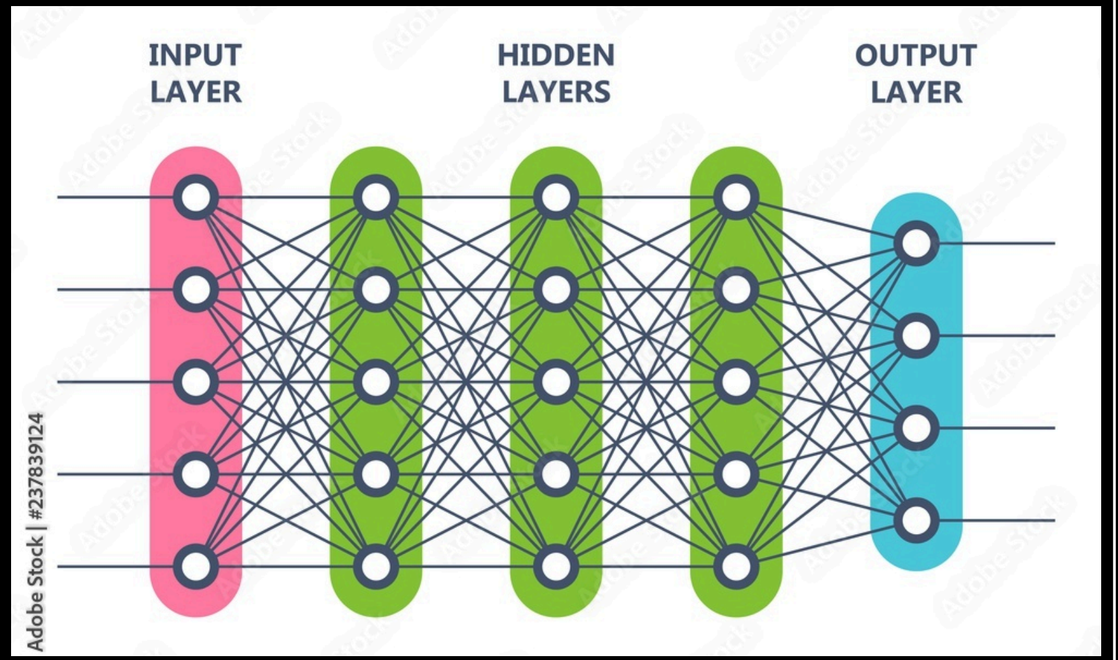
Un RNA est généralement composé de plusieurs couches de neurones :

Couche d'entrée : Représente les données d'entrée.

Couches cachées : Effectuent des transformations non linéaires sur les entrées.

Couche de sortie : Produit la sortie finale du réseau.

Chaque couche ll contient un certain nombre de neurones. La sortie d'une couche devient l'entrée de la couche suivante.



Propagation avant (Forward Propagation)

Fonction de sommation : calcule une combinaison linéaire des entrées en effectuant la somme de chaque valeur d'entrée multipliée par son poids associé :

$$z_j^{[l]} = \sum_i w_{ji}^{[l]} A_i^{[l-1]} + b_j^{[l]}$$

- $A_i^{[l-1]}$: Sortie du neurone i de la couche précédente $l-1$ (**activation**).
- $W_{ji}^{[l]}$: Poids reliant le neurone i de la couche $l-1$ au neurone j de la couche l.
- $b_j^{[l]}$: Biais du neurone j dans la couche l.

Ensuite, on applique une fonction d'activation f pour obtenir les sorties de la couche :

$$A^{[l]} = f(Z^{[l]})$$

Fonction d'activation

Une fonction d’activation introduit de la non-linéarité dans un réseau de neurones, permettant au modèle d’apprendre des relations complexes. Elle aide également à la convergence du modèle et à la propagation du gradient.

Nom	Formule	Utilisation
Sigmoïde	$f(x) = \frac{1}{1+e^{-x}}$	Classification binaire
Tanh (Tangente Hyperbolique)	$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	Réseaux de neurones récurrents (RNN)
ReLU (Rectified Linear Unit)	$f(x) = \max(0, x)$	Réseaux profonds (CNN, MLP)
Softmax	$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	Classification multi-classes

Fonction de coût (Loss Function)

La fonction de coût (Loss Function) joue un rôle central en apprentissage automatique, que ce soit pour des tâches de régression ou de classification. Elle mesure à quel point les prédictions du modèle sont éloignées des valeurs réelles (ou des labels). L'objectif principal du modèle est de minimiser cette fonction de coût, ce qui permet d'améliorer la précision des prédictions

Fonction de coût (Loss Function)

Nom	Formule	Utilisation
Erreur Quadratique Moyenne (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Régression
Erreur Logarithmique (Log Loss)	$-\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$	Classification binaire
Hinge Loss	$\sum_{i=1}^{n}\max(0, 1-y_i\hat{y}_i)$	Classification SVM
Kullback-Leibler Divergence (KL-Divergence)	$\sum_i P(i)\log\frac{P(i)}{Q(i)}$	Apprentissage probabiliste
Cross-Entropy Loss (CE)	$-\sum_i y_i \log(\hat{y}_i)$	Classification multi-classes

Rétropropagation (Backpropagation)

La rétropropagation est un algorithme essentiel pour entraîner un réseau de neurones. Son objectif est d'ajuster les poids du réseau afin de minimiser l'erreur entre les sorties prédîtes et les sorties réelles . voici les étapes de cette partie :

1. Calculer le gradient de la fonction de coût par rapport à la sortie :

$$\frac{\partial L}{\partial \hat{y}}$$

2. Propager l'erreur à travers les couches en utilisant la règle de la chaîne :

$$\frac{\partial L}{\partial w_{ji}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot \frac{\partial a_j^{(l)}}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_{ji}^{(l)}}$$

3. Mettre à jour les poids et les biais en utilisant la descente de gradient :

$$w_{ji}^{(l)} \leftarrow w_{ji}^{(l)} - \eta \frac{\partial L}{\partial w_{ji}^{(l)}}$$

où η est le taux d'apprentissage.

Dans la phase de mise à jour des poids, plusieurs variantes de la descente de gradient peuvent être utilisées pour améliorer la convergence et la stabilité de l'apprentissage. Parmi les approches les plus courantes, on retrouve :

- Descente de gradient stochastique (SGD)
- Descente de gradient par mini-lots (Mini-batch GD)
- Momentum
- RMSprop
- Adam (Adaptive Moment Estimation)

Modélisation finale

❑ **Données :**

- Soit $A = \{(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\}$ l'ensemble des données d'apprentissage.
- Chaque $X_i \in \mathbb{R}^d$ est un vecteur d'entrée, et $y_i \in \mathbb{R}$ est la sortie cible (dans le cas d'une régression).

❑ **Variables :**

- $W^{(l)}$: matrice des poids de la couche l, de taille $d_{l+1} \times d_l$.
- $b^{(l)} \in \mathbb{R}^{d_{l+1}}$: vecteur des biais de la couche l.
- $Z^{(l)} = W^{(l)} A^{(l-1)} + b^{(l)}$: sortie linéaire de la couche l.
- $A^{(l)} = f^{(l)}(Z^{(l)})$: sortie activée de la couche l, où $f^{(l)}$ est la fonction d'activation (ReLU, sigmoid, etc.).

❑ **Objectif :**

Minimiser la fonction de coût (par exemple, l'erreur quadratique moyenne pour la régression) :

$$\min_{W,b} \mathcal{L}(W,b) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

avec

$$\hat{y}_i = A^{(L)}(X_i)$$

où L est l'indice de la dernière couche du réseau.

❑ **Contraintes :**

- $W^{(l)} \in \mathbb{R}^{d_{l+1}\times d_l}$, $b^{(l)} \in \mathbb{R}^{d_{l+1}}$
- Fonctions d'activation $f^{(l)}$ bien définies (ReLU, tanh, etc.)
- Propagation avant :

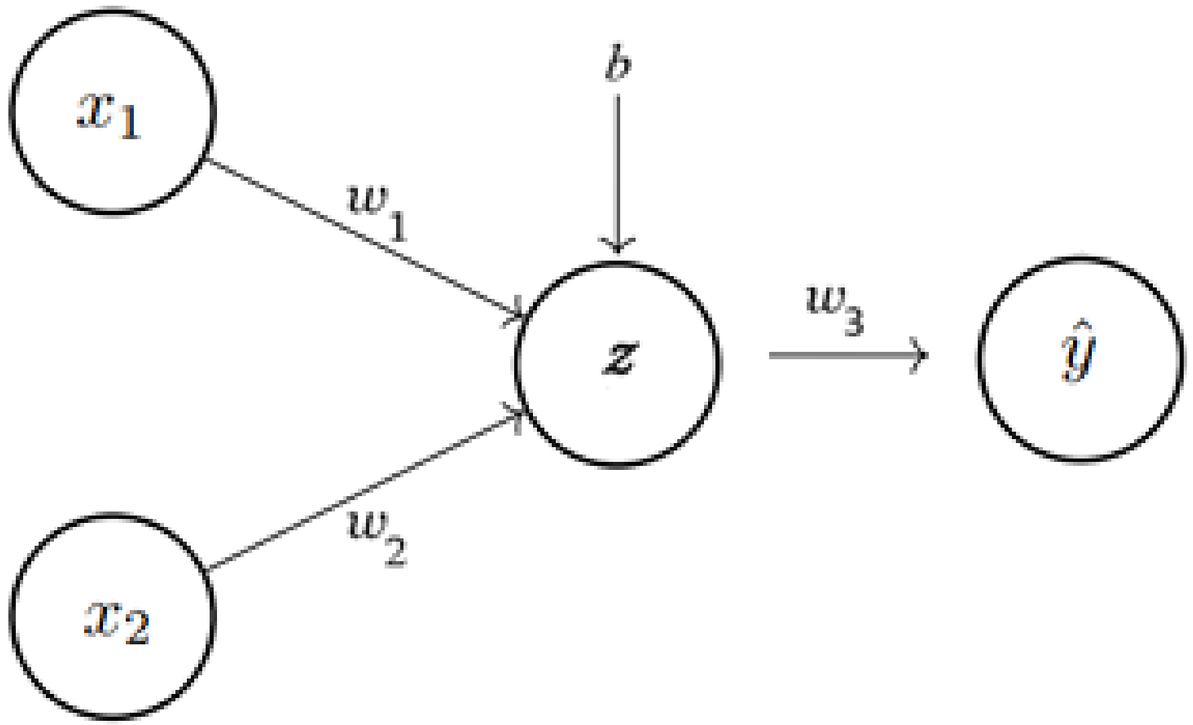
$$A^{(0)} = X_i, \quad Z^{(l)} = W^{(l)}A^{(l-1)} + b^{(l)}, \quad A^{(l)} = f^{(l)}(Z^{(l)})$$

Application

Soit un réseau de neurones avec les valeurs suivantes :

- x1=0.5 , x2=0.8
- w1=0.2 ,w2 = 0.6, w3=1.0
- b=1

taux d’apprentissage=0.1



La fonction d’activation est sigmoïde.

Appliquer le modèle mathématique du Réseau de Neurones Artificiel (RNA) .

Applications et Défis

Applications :

- Reconnaissance d’image
- Traitement du langage naturel
- Systèmes de recommandation

Défis :

- Modélisation complexe pour les grands réseaux
- Interprétation difficile
- Sensible aux hyperparamètres
- Moins utilisée directement en pratique

Conclusion

Les réseaux de neurones artificiels, par leur puissance mathématique et leur capacité d’adaptation, transforment des données en savoirs profonds. Leur architecture soignée et leurs méthodes d’apprentissage ouvrent des horizons infinis dans l’intelligence artificielle, de la vision par ordinateur à la robotique.

UNIVERSITÉ ABDELMALEK ESSAÂDI

Merci !

QUESTIONS???

Présentée par Soukaina El Messaoui