

Introduction au Deep LEarning

Réseaux récurrents

Olivier Schwander <olivier.schwander@sorbonne-universite.fr>

2025-2026

Exercice 1 - Réseaux récurrents simples

Un réseau récurrent simple (RNN) traite une séquence d'entrées x_t en mettant à jour un état caché h_t à chaque pas de temps t :

$$h_t = \tanh(Wx_t + Uh_{t-1} + b_h)$$

$$y_t = Vh_t + b_o$$

où

- $x_t \in \mathbb{R}^{d_{in}}$ est l'entrée au pas de temps t
- $h_t \in \mathbb{R}^{d_h}$ est l'état caché au pas de temps t
- $y_t \in \mathbb{R}^{d_{out}}$ est la sortie au pas de temps t
- $W \in \mathbb{R}^{d_h \times d_{in}}, U \in \mathbb{R}^{d_h \times d_h}, V \in \mathbb{R}^{d_{out} \times d_h}$ sont les matrices de poids
- $b_h \in \mathbb{R}^{d_h}, b_o \in \mathbb{R}^{d_{out}}$ sont les biais

La fonction d'activation $\tanh$ assure que l'état caché reste dans l'intervalle $[-1, 1]$.

Question 1

On voit Figure 1 le graphe de dépendances d'un RNN à l'instant t . Développer ce graphe pour une séquence $\{x_1, x_2, \dots, x_t\}$.

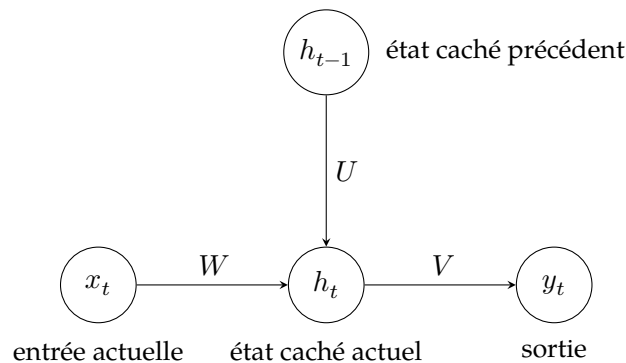


FIGURE 1 : Représentation d'un RNN

Question 2

Pour une longueur de séquence d'entrée fixée, on veut réécrire ce RNN sous la forme de couches denses. Quelle contrainte doit-on rajouter sur les matrices de paramètres ?

Question 3

Pour une séquence d'entrée très longue, montrer qu'on est confronté au problème du vanishing gradient.

Question 4

Commenter la non-linéarité $\tanh$ utilisée pour les RNN.

Question 5

On peut bien sûr empiler plusieurs couches récurrentes. Quelle interprétation peut-on en faire ?

Question 6

Expliquer comment utiliser un RNN en génération puis en classification.

Exercice 2 - Long Short Term Memory (LSTM)

Question 7

L'unité LSTM introduit des portes (gates) pour contrôler le flux d'information.

Portes :

— **Porte d'oubli** (forget gate) f_t :

$$f_t = \sigma(W_f x_t + U_f c_{t-1} + b_f)$$

— **Porte d'entrée** (input gate) i_t :

$$i_t = \sigma(W_i x_t + U_i c_{t-1} + b_i)$$

— **Porte de sortie** (output gate) o_t :

$$o_t = \sigma(W_o x_t + U_o c_{t-1} + b_o)$$

États :

— **État de cellule** c_t :

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_t + b_c)$$

— **État caché** h_t :

$$h_t = o_t \odot \tanh(c_t)$$

Question 8

Analyser l'influence des valeurs des portes d'oubli et d'entrée sur la valeur de la cellule.

Question 9

Pourquoi l'équation de mise à jour de la cellule permet-elle d'éviter le problème du vanishing gradient ?

Question 10

Pourquoi la cellule permet-elle de mémoriser des informations sur le long terme, contrairement à l'état caché?

Question 11

Quel état utiliser pour la classification ou pour empiler des couches LSTM?

Exercice 3 - *Application au texte*

L'application à un texte nécessite de d'avoir en entrée des couches récurrentes une représentation des mots.

On utilise une matrice d'embeddings de taille nombre de mots $\times d$ avec d un hyper-paramètre. La ligne i de la matrice contient l'embedding du mot i . Lors de la passe forward, on sélectionne une ligne de la matrice en fonction du mot considéré.

Question 12

Pourquoi la représentation one-hot-encoding est-il inadaptée comme entrée?

Question 13

Montrer que la sélection d'une ligne avec un indice peut s'écrire comme un produit matrice-vecteur avec un vecteur one-hot en entrée.

Quel est le souci de cet approche explicite?

Question 14

Comment faire pour la rétropropagation?