

Réponses aux questions

IFT780 - tp3

mnist_autoencoders.ipynb

Question 1 :

- Quelle est la relation entre l'architecture de l'encodeur et du décodeur ?
- Pourquoi la fonction d'activation en sortie est-elle une **sigmoïde** ?
- Quelle est la taille de l'espace latent du modèle?

Question 2 :

- Quel est l'utilité du dataloader?
- Que fait la propagation avant considérant qu'elle utilise 2 réseaux de neurones? Quelles sont les entrées et sorties de cette fonction?

Question 3 :

Pourquoi pensez-vous que l'espace latent de deux dimensions donne des reconstructions plus floues et moins précises ?

Question 4:

- La cellule précédente multiplie par 2 la taille de l'espace latent lors de l'instanciation de l'encodeur. Pourquoi ?
- Dans la prochaine cellule, une certaine opération, une astuce, est effectuée afin d'échantillonner l'espace latent. Quelle est cette astuce et pourquoi est-elle utilisée?
- Quelle est la taille de notre espace latent ?

Question 5:

Quelle différence remarquez-vous entre les deux espaces latents ?

cardiac_mri_autoencoders.ipynb

Question 1 :

- Combien de neurones contient l'encodeur ?
- Combien de paramètres contient l'encodeur ?
- Quelle est la taille de l'espace latent ?

Question 2 :

En quoi consiste l'opération ConvTranspose2d? Pourquoi est-elle nécessaire?

Question 3:

- Quelles sont les différences entre les fonctions *vae_forward_pass* de ce notebook et celui de l'autoencodeur sur MNIST ?

Question 4:

- Réentraînez l'autoencodeur avec un espace latent à 2 dimensions. Comme dans le cas de MNIST, les reconstructions sont de bien moins bonne qualité. En quoi selon vous ces résultats sont de moins bonne qualité et pourquoi ne sont-ils pas flous comme pour MNIST?

rnn_image_captioning.ipynb

Question 1 :

- À quoi servent les jetons spéciaux [START], [END], [PAD], [UNK] ? Donnez une réponse par jeton.

Question 2:

- Pourquoi utiliser un modèle pré-entraîné ? Pourquoi ne pas entraîner aussi l'encodeur ?
- Expliquez dans vos propres mots quelle est la représentation de l'image en sortie de l'encodeur. Considérez la taille du tenseur, ce qu'il représente, et l'architecture Resnet50.

Question 3:

- Quel est l'utilité de la "embedding layer"?
- Que représente la sortie du modèle dans la fonction forward? Et dans la fonction predict?
- Quelles sont les différences entre l'entrée du LSTM dans la fonction forward et la fonction predict? Quels sont les avantages d'avoir ces différences ?

Question 4

- Pourquoi utiliser l'entropie-croisée comme fonction de perte ? Pourquoi pas une fonction de perte L2, par exemple?

Questions 5

- Les descriptions générées à la cellule précédente devraient parfois manquer de précision, inclure des éléments de trop ou en omettre ("a group of cows standing ..." versus "a cow standing ...", par exemple). Pourquoi croyez-vous que cela se produit ? Prenez en compte l'architecture des deux modèles et les tenseurs échangés entre ceux-ci. Indice: ce n'est pas dû à l'entraînement ou parce que le modèle n'a pas convergé ou par manque de données. Si vous êtes courageux.euses, vous pouvez entraîner le modèle plus longtemps pour vous en convaincre.
- Pour cette implémentation, nous avons utilisé un ensemble de jetons correspondant grosso modo à chaque mot dans l'ensemble de toutes les descriptions. Combien de jetons font partie du vocabulaire du modèle ?
- Quel aurait été l'impact sur le modèle, sur son nombre de paramètres et sur la difficulté du problème si nous avions choisi d'avoir plutôt un jeton par caractère ?