

Documents multimédia : description et recherche automatique

M2 GI, examen du 13 avril 2022, 2 heures, calculatrice et documents autorisés.

Les réponses doivent être justifiées de manière concise.

Partie 1. Apprentissage classique

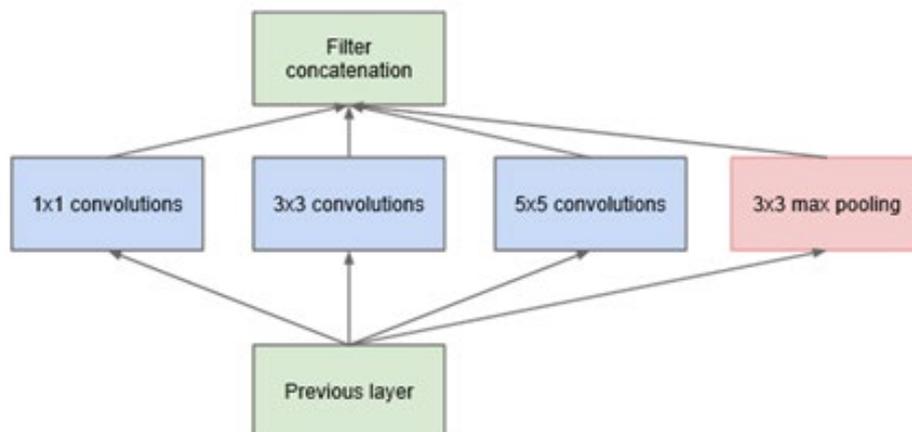
Question 1.1 : À quoi sert une normalisation en probabilité ? Comment peut-on la réaliser ?

Question 1.2 : Pourquoi utilise-t-on un ensemble de validation en apprentissage par machine ?

Question 1.3 : Quelles méthodes vues dans le cours peut-on utiliser pour implémenter un classificateur sur des données non linéairement séparables ?

Partie 2. Module « inception »

La figure suivante donne l'architecture d'un module « inception » simple ou « naïf » :



Question 2.1 : Quel est l'intérêt de ce type de module par rapport à une couche de convolution simple ?

L'entrée venant de la couche précédente est composée de 24 plans (ou filtres ou « feature maps » ou « channels »). Chaque bloc de convolution sort sur 8 plans.

Question 2.2 : Quel est le nombre total de plans à la sortie du module « inception » ? (La concaténation se fait dans la dimension des plans.)

Les images d'entrée sont de taille 56×56.

Question 2.3 : Quelle est la taille des images de sortie après chacune des trois convolutions si on ne fait pas de « padding » ?

Question 2.4 : Quelles valeurs de « padding » faut-il pour chacune d'elle si on veut pouvoir concaténer (empiler) les sorties de toutes ces convolutions ?

Le « pooling » impliqué ici est fait avec un « stride » de 1, c'est à dire qu'il est calculé à toutes les positions, un peu comme dans une convolution 3×3. Toutes les convolutions incluent des biais.

Question 2.5 : Combien de paramètres y a-t-il dans chaque bloc de convolution ? Dans le module « inception » dans son ensemble ?

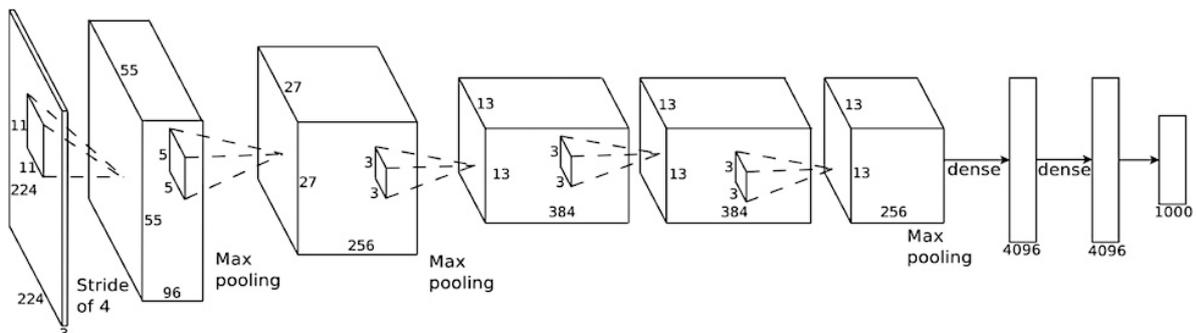
Question 2.6 : Combien d'opérations flottantes sont-elles effectuées lors du passage d'une image dans chaque bloc de convolution ? Dans le module « inception » dans son ensemble ?

Question 2.7 : Dans quelle architecture classique trouve-t-on des modules « inception » ?

Partie 3. Réseau « AlexNet »

La figure ci-dessous représente de manière simplifiée une variante du réseau « AlexNet » qui a remporté le « ImageNet Large Scale Visual Recognition Challenge » (ILSVRC) en 2012. Ce réseau prend en entrée des images normalisées en taille (224×224). Il donne directement en sortie des scores de détection pour chacun des 1000 concepts cibles. Les éléments suivants ne sont pas détaillés sur la figure :

- les couches neuronales contiennent toutes une partie linéaire (affine en fait car des biais sont utilisés) et une partie non linéaire (la fonction d'activation) ;
- les sixième et septième couches neuronales contiennent en plus en entrée une étape dite de « dropout » ;
- une étape finale de « softmax » est ajoutée à la sortie du réseau.



Question 3.1 : Combien de couches « neuronales » ce réseau contient-il ? Quels types de couches neuronales contient-il et combien en contient-il de chaque type ?

Question 3.2 : Combien de paramètres contient la deuxième couche neuronale ?

Question 3.3 : Combien de connexions contient la huitième couche neuronale ?

Question 3.4 : En comptant 1 pour chaque étape linéaire, 1 pour chaque fonction d'activation, 1 pour chaque étape de « pooling », 1 pour chaque étape de « dropout », et 1 pour l'étape finale de « softmax », de combien d'étapes au total le réseau est-il composé~?

Question 3.5 : Comment la transition entre les couches de convolution et les couches complètement connectées est-elle effectuée ?

Question 3.6 : Quel est le but des étapes de « dropout » (non représentées sur la figure) et comment sont-elles implémentées ?

Question 3.7 : À quoi sert l'étape finale de « softmax » (non représentée sur la figure) et comment est-elle implémentée ?

Question 3.8 : Dans quel cas l'étape finale de « softmax » est-elle adaptée et dans quel cas ne l'est-elle pas ?

Partie 4. Apprentissage profond et recherche d'instance

Question 4.1 : Quelle technique a permis l'entraînement de réseaux très profonds « very deep » (environ 20 couches) ?

Question 4.2 : Quelle technique a permis l'entraînement de réseaux ultra profonds « ultra deep » (plus de 100 couches) ?

Question 4.3 : À quoi servent les réseaux dits « Siamois » ?

Question 4.4 : Quelle fonction de coût leur est habituellement associée ?