

Polytech' Grenoble – Info 4
Examen Accès et Recherche d'Information – 12 avril 2022 – 2 heures

Note importante : les parties I et II sont à rédiger sur des copies séparées.

Indication : 5 minutes par question en moyenne, nous attendons des réponses concises.

Partie I

Exercice 1 – Généralité sur les modèles et systèmes de recherche d'information

Expliquez pourquoi, en vous basant sur un exemple si nécessaire, l'utilisation de tableaux de rappel/précision (normalisés ou non), sont nécessaires si on veut intégrer le fait que les réponses des systèmes sont triées. (5 lignes max)

Exercice 2 : Traitement de synonymes par un SRI vectoriel

Considérons un corpus formé des 2 documents suivants :

$D_1 = \text{"auto électrique jolie voiture"}$

$D_2 = \text{"véhicules verts bleus"}$

Question 1 : En considérant l'utilisation l'algorithme de Porter ayant les règles suivantes :

1. $s \rightarrow /$
2. $iqu \rightarrow ic$
3. $ien \rightarrow /$
4. $ion \rightarrow /$
5. $e \rightarrow /$

indiquer les termes qui indexent les documents (en décrivant les règles de réécritures utilisées pour chaque mot traité). En déduite le vocabulaire qui indexe ces documents, trié par ordre alphabétique de ces termes.

Question 2 : Donner les valeurs de p_{tf} de chaque terme t_i dans chaque document, ainsi que le p_{df} de chaque terme.

Question 3 : En déduire les vecteurs de ces deux documents, en supposant une normalisation $n_d=1$ (comme en cours et en TD).

Question 4 : Posons une requête q , "voitures". Donner le terme obtenu par troncature (cf. question 2), son p_{tf} et son p_{df} (comme en TP).

Question 5 : En déduire le vecteur de la requête.

Question 6 : Calculer la similarité (RSV) entre le vecteur obtenu en question 5 et les vecteurs deux documents obtenus en question 3. En déduire la réponse à la requête q .

Question 7 : Commenter de manière courte le résultat de la requête q .

Question 8 : On va supposer que les mots "voiture", "auto" et "véhicule" sont synonymes. Une manière d'utiliser cette information est de faire de "l'expansion de requête" en ajoutant

tous les synonymes des termes qui apparaissent dans la requête. Donner la requête (chaîne de caractères) de la question 5 étendue, appelée q' .

Question 9 : Calculer le vecteur qui correspond à q' (question similaire aux questions 5 et 6).

Question 10 : Calculer la similarité (RSV) entre la requête q' les deux documents. En déduire la réponse à la requête q' .

Question 11 : Commenter les différences avec la question 7. A votre avis, est-ce que l'utilisation de l'expansion de requête peut favoriser le rappel, la précision, ou les deux ?

Partie 2

Exercice 1. Apprentissage classique

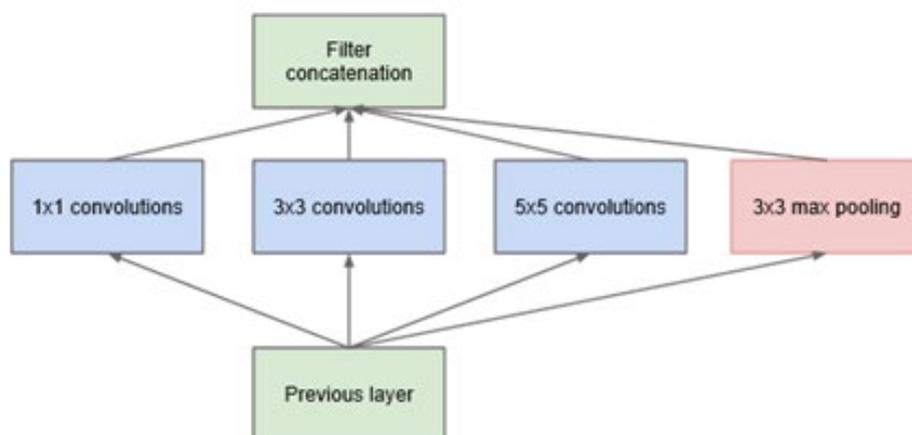
Question 1.1 : Quelles méthodes vues dans le cours peut-on utiliser pour implémenter un classificateur sur des données non linéairement séparables ?

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

Question 1.3 : À quoi sert une normalisation en probabilité ? Comment peut-on la réaliser ?

Exercice 2. Apprentissage profond pour la classification

La figure suivant 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 16 plans ou filtres (« features » ou « channels »). Chacun des blocs de convolution sort sur 8 plans.

Question 2.2 : Quel est le nombre total de plans à la sortie du module « inception » ?

Les images d'entrée sont de taille 32×32 .

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 .

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 sont 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 » ?

Exercice 3. Apprentissage profond

Question 3.1 : Quelle technique a permis l'entraînement de réseaux très profonds « very deep » (environ 20 couches) ?

Question 3.2 : Quelle technique a permis l'entraînement de réseaux ultra profonds « ultra deep » (plus de 100 couches) ?

Question 3.3 : À quoi servent les réseaux dits « Siamois » ?

Question 3.4 : Quelle fonction de coût leur est habituellement associée ?