

Documents multimédia : description et recherche automatique

M2 GI, examen du 19 avril 2023, 2 heures, calculatrice et documents autorisés.

Les réponses doivent être justifiées de manière concise.

Partie 1. Apprentissage classique, SVM

Dans la famille des classificateurs SVM, on considère (a) les SVM de base, (b) les SVM avec « marge souple » (soft margin) et (c) les SVM à noyaux avec fonctions à base radiale (RBF kernels).

Question 1.1 : Laquelle ou lesquelles sont adaptées pour des données avec des classes linéairement séparables ?

Question 1.2 : Laquelle ou lesquelles sont adaptées pour des données avec des classes « presque » linéairement séparables ?

Question 1.3 : Laquelle ou lesquelles sont adaptées pour des données avec des classes non linéairement séparables ?

Question 1.4 : Dans quel cas des données correspondent-elles à des classes non linéairement séparables ?

Partie 2. Descripteurs classiques, histogrammes de couleur

Question 2.1 : Quelle est la dimension d'un descripteur d'image basé sur un histogramme de couleurs tridimensionnel dans l'espace LAB avec respectivement 8, 5 et 5 « bins » pour les composantes L, A et B ?

Question 2.2 : Quelles distances peuvent être utilisées pour évaluer la similarité de ces histogrammes dans le cadre d'une recherche par l'exemple ?

Question 2.3 : Comment un histogramme RGB normalisé est-il modifié si on double la taille de l'image ?

Question 2.4 : Comment un histogramme RGB normalisé est-il modifié si on transforme une image par symétrie gauche-droite ?

Question 2.5 : En quoi consiste la normalisation d'un histogramme ?

Partie 3. Apprentissage profond pour la classification

On considère un réseau à propagation avant qui prend en entrée des images RGB de taille 64×64 et qui applique successivement les opérateurs suivants :

- un module de 64 filtres de convolution de taille 5×5 sans « padding » ;
- une rectification linéaire « ReLU »
- une réduction de taille par « max pooling » par blocs de taille 3×3
- un module de 32 filtres de convolution de taille 5×5 sans « padding » ;
- une rectification linéaire « ReLU »
- une réduction de taille par « max pooling » par blocs de taille 2×2
- un module linéaire avec 256 sorties
- une rectification linéaire « ReLU »
- un module linéaire avec 20 sorties
- une couche de « softmax »

Question 3.1 : Combien de couches neuronales ce réseau contient-il ?

Question 3.2 : Quel est le nombre total de plans à la sortie du premier et du second module de convolution ?

Question 3.3 : Quelle est la taille des images de sortie après chacune des deux modules de convolution ?

Question 3.4 : Combien de paramètres « apprenables » contiennent chacune des fonctions ?

Question 3.5 : Combien d'opérations sont effectuées dans le second module de convolution pour une image ?

Question 3.6 : Combien d'opérations sont effectuées dans le module linéaire pour une image ?

Question 3.7 : Quel est le but de la couche de « softmax » ?

Question 3.8 : Pourquoi ne met-on pas de rectification linéaire « ReLU » après le dernier module linéaire

Partie 4. Transfert d'apprentissage

Question 4.1 : Donnez un exemple de transfert d'apprentissage.

Question 4.2 : Qu'est-ce qu'un « backbone » dans ce contexte ?