

Licence d'Informatique 2

Analyse de Données Utilisateur (C5-160412)

TP 5 – Apprentissage Supervisé (Prédiction)

Carl FRÉLICOT – Dpt Info / Lab MIA

1. Données Quantitatives de l'Exercice 2 du TD 5

- (a) Chargez les données de la feuille **Learning Set 1** et de la feuille **Unknown** du fichier **TD5-2d.xlsx** séparément.
- (b) Trouvez le moyen de visualiser simultanément les deux nuages correspondants.
- (HW) Comment sélectionner un individu inconnu particulier (en l'occurrence x , y ou z) dont on souhaite trouver les *plus proches voisins* dans le tableau d'apprentissage ? Connectez le composant adéquat.
- (HW) À l'aide du composant **Neighbors**, retrouvez les *plus proches voisins* (au sens de la distance de Manhattan) des inconnus.
- (c) Comprenez comment utiliser les composants **kNN** et **Predictions** afin d'appliquer la règle du 1-PPV sur l'ensemble des inconnus, puis effectuez le classement. Vous expliquerez les colonnes de la table résultat.
- (d) De même avec la règle des 3-PPV.
- (HW) Chez vous, vous serez en mesure de vérifier les calculs du (HW) de la feuille de TD 5.
- (e) Vérifiez le problème lié à l'utilisation, sur ces données, des règles des 2-PPV et des 4-PPV.
- (HW) Hors séance, vous pourrez exécuter ce *canvas* en remplaçant la feuille **Learning Set 1** par la feuille **Learning Set 2**.



2. Données Quantitatives de l'Exercice 1 du TD 5

- (a) Dans un nouveau *canvas*, chargez les données de la feuille 1 nouveau fichier **terminale.xlsx** contenant des notes d'élèves et l'avis pour le baccalauréat.
- (b) Retrouvez les *facteurs* et les *vecteurs directeurs* des axes qui les définissent à l'aide du composant **Linear Projection**, puis visualisez le nuage de points dans le (premier) *plan discriminant*.
- (c) Chargez séparément les données de la feuille 2, puis rappelez ce qu'il faudrait faire pour prédire l'avis de ces autres élèves.
- (d) Projetez ces données dans l'espace discriminant à l'aide du composant **Transform**, puis vérifiez les valeurs trouvées en TD.
- (e) Faites le nécessaire pour visualiser simultanément les deux nuages dans cet espace.

Orange ne permet pas de calculer de manière simple les centres des groupes dans l'espace discriminant afin de prédire... On vous propose d'utiliser les PPV.

- (f) Classez les nouveaux élèves à l'aide de la règle du 1-PPV, et la distance euclidienne usuelle.
- (g) À partir de quel nombre K de voisins obtient-on des prédictions non satisfaisantes ?

3. Données Réelles – Voyage, voyage...

- (a) Décompressez le fichier **cities.zip**, ouvrez un troisième *canvas* et importez les images qu'il contient à l'aide du composant approprié.
- (b) Connectez un **Data Table** afin de savoir ce qu'il en est.
- (c) Trouvez le moyen de visualiser les images.
- (d) Un réseau de neurones profond a été codé dans le composant **Image Embedding** afin d'extraire 1 000 caractéristiques numériques de chaque image quelle que soit sa taille. Connectez ce composant et regardez le tableau de données en résultant.
- (e) Projetez les images ainsi codées dans le premier plan discriminant.
- (f) Connectez les composants permettant de *reclasser* la projection des images par la règle des K -PPV.
- (g) On souhaite visualiser ce qu'on appelle la *matrice de confusion*, c'est-à-dire le tableau croisant les groupes d'origine et les groupes prédits. Que vaut-elle pour les 1-PPV ?
- (h) Testez plusieurs distances et valeurs de K de sorte de sélectionner la meilleure méthode. Vous afficherez les mauvaises prédictions, éventuellement les images.
- (i) Pourquoi *reclasser* n'est pas une bonne manière d'évaluer une méthode de classification supervisée ?

4. Données Réelles – Mangez 5 fruits et légumes par jour !

- (a) Chargez les données du fichier **fruitsvegs.xlsx** et **fruitsvegsnews.xlsx** dans le premier *canvas*, puis regardez ce qu'ils contiennent.
- (b) À votre avis, pourquoi ne peut-on pas utiliser le deuxième *canvas* ?
- (c) Utilisez le premier, et classez les nouveaux individus à l'aide de la règle des K -PPV. Vous pourrez tester plusieurs valeurs de K , plusieurs distances.