

Interprétation des résultats HRAP

Ce document explique comment exécuter le programme d'interprétations des résultats

Mode d'emploi

Dans le fichier contenant l'exécutable, double-cliquez sur l'application job_assessment.exe.
Une fenêtre de type console va apparaître et un menu va s'afficher (met 1-2 secondes)

```
----- MENU -----
1. Predict
2. Accuracy
3. Exit
-----
Enter your choice [1-3]:
```

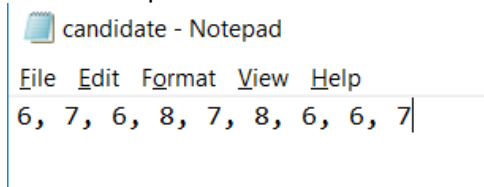
Choisissez le mode voulu :

Le mode « Predict » permet de prédire un profil donné à partir d'une base de données (fichier profil.csv)
Le mode « Accuracy » permet d'observer la performance du classifieur donné sur un set d'entraînement et un set d'apprentissage donné.

Mode Predict

Le mode « Predict » permet de prédire un profil donné qui est actuellement sous le nom « candidate.csv »

Celui-ci se présente comme ci-dessous :



Il contient le vecteur des compétences suivantes :

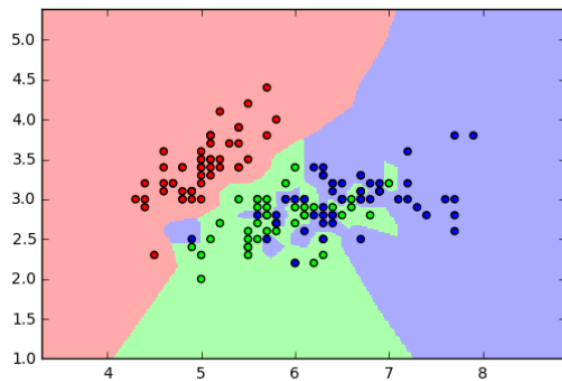
Leadership	Sociabilité	Contrôle émotionnel	Atteinte des objectifs	Avant- vente	Pilotage Suivi	Relation client	Gestion équipe	Reporting
------------	-------------	------------------------	---------------------------	-----------------	-------------------	--------------------	-------------------	-----------

Choisissez un classifieur

```
----- Choose a classifier -----  
1. Nearest Neighbors  
2. Logistic Regression  
3. Exit  
-----  
Enter your choice [1-3]:
```

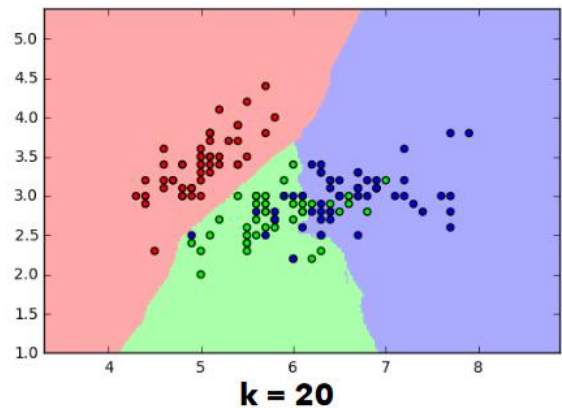
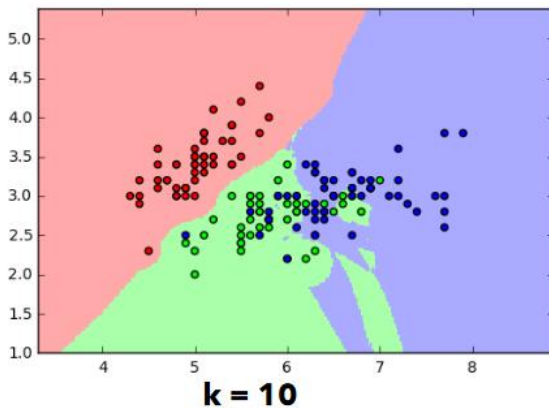
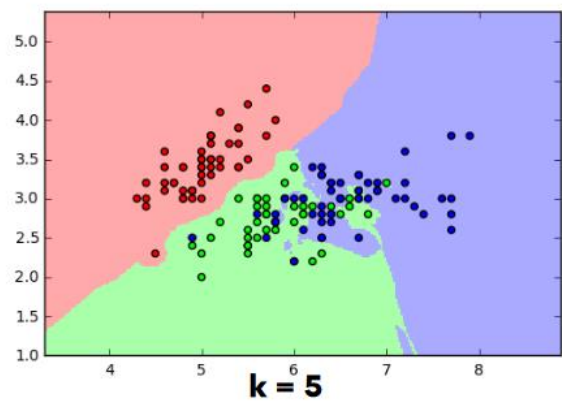
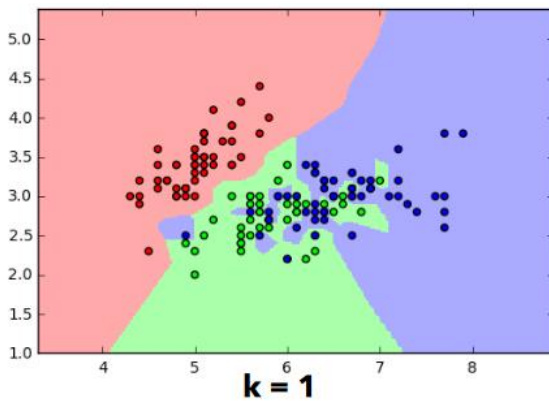
Sélectionnez ensuite le modèle voulu :

1. Méthode des k plus proches voisins



1-plus proche voisin

Découpage de l'espace en
« zone d'influence » de
chaque point de l'ensemble
d'entraînement



Entrez le paramètre désiré pour k :

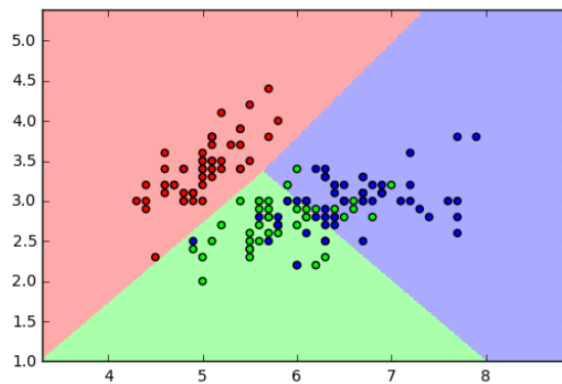
Enter number of neighbors [1-9]:

Une fonction est actuellement en cours de développement pour sélectionner le meilleur paramètre k entre 1 et 9

Exemple de résultat avec la méthode des k plus proches voisins:

```
2018-01-12 17:37:43,604 - INFO - Use kNN classifier with k= 5
2018-01-12 17:37:43,604 - INFO - Train set size is (20, 9)
2018-01-12 17:37:43,604 - INFO - Training done in 0.000s
Prediction for the given profile
['DP']
```

2. Régression logistique



- **Avantage** : modèle simple et stable
- **Inconvénient** : les données doivent être linéairement séparables, modèle trop simple

Comparaison k-PPV et régression logistique

k-PPV

- **Avantage** : la frontière de classification peut être très complexe, sa forme n'est pas limitée
- **Inconvénient** : la frontière est instable, elle dépend fortement des exemples

Régression logistique

- **Avantage** : la frontière de classification est stable, elle ne varie pas beaucoup en fonction des exemples
- **Inconvénient** : la frontière est limitée à des droites (hyperplans)

Exemple de résultat avec la régression logistique :

```
2018-01-12 17:46:59,537 - INFO - Use logistic_regression classifier
2018-01-12 17:46:59,538 - INFO - Train set size is (20, 9)
2018-01-12 17:46:59,541 - INFO - Training done in 0.001s
Prediction for the given profile
['DP senior']
```

Mode Accuracy

Ce mode permet d'évaluer la pertinence d'un classifieur donné avec la base de données choisie en entrée, plus le nombre de données est important, meilleure est la précision de ce classifieur. Ce mode est peu pertinent pour un jeu pauvre en données (<100)

Sachez qu'à chaque test effectué, la base de données est séparée aléatoirement en 2 «sets»: un set d'entraînement (80% des données) et un set d'apprentissage (20% des données).

Choisissez un classifieur à évaluer :

```
----- Choose a classifier -----  
1. Nearest Neighbors  
2. Logistic Regression  
3. Nearest Neighbors Logistic Resresion  
4. Exit
```

1. Méthode des k plus proches voisins

Cette fonction mesure la précision de la méthode des k plus proches voisins, avec k donné en entrée, sur un set d'entraînement aléatoire.

```
Enter your choice [1-4]: 1  
Enter a number of neighbors [1-9]: 2  
2018-01-12 18:02:30,969 - INFO - Use kNN classifier with k= 2  
2018-01-12 18:02:30,972 - INFO - Train set size is (16, 9)  
2018-01-12 18:02:30,972 - INFO - Test set size is (4, 9)
```

Vérifiez que le nombre k donné en entrée soit inférieur au nombre de données différentes d'une classe.

Si le programme se ferme c'est qu'une erreur sur la valeur de k est apparue :

```
ValueError: Expected n_neighbors <= n_samples, but n_samples = 3, n_neighbors = 8
```

2. Régression logistique

Cette fonction mesure la précision de la régression logistique sur un set d'entraînement aléatoire.

3. Méthode des k plus proches voisins et Régression logistique

Cette fonction mesure la précision de la méthode des k plus proches voisins (avec le meilleur k entre 1 et 25) et la précision de la régression logistique, sur un même set d'entraînement aléatoire.

Si le nombre d'échantillons total pour chaque classe est inférieur à 25, cette fonction ne fonctionnera pas.

```

----- Choose a method -----
1. Learning curve
2. Testing curve
3. No curve

```

1. Courbe d'apprentissage

Affiche l'impact de la taille du set d'apprentissage sur la précision du classifieur

Si le nombre d'échantillons total est assez bas, cette fonction n'est pas pertinente et peut entraîner des erreurs.

2. Courbe d'entraînement

Affiche l'impact de la taille du set d'entraînement sur la précision du classifieur

Si le nombre d'échantillons total est assez bas, cette fonction n'est pas pertinente et peut entraîner des erreurs.

3. Aucune courbe

Affiche le score de précision, pour un classifieur donné, sur le set d'apprentissage.

Affiche le score de précision, pour un classifieur donné, sur le set d'entraînement.

Affiche les statistiques pour le test d'entraînement.

Exemple de résultat :

```

Knn :
2018-01-29 13:24:31,680 - INFO - Training done in 0.000s
2018-01-29 13:24:31,680 - INFO - Testing KNN
2018-01-29 13:24:31,680 - INFO - Testing done in 0.000s
Score on training : 0.883721
Score on testing : 0.818182

```

	precision	recall	f1-score	support
Architecte	0.00	0.00	0.00	2
CP	0.75	1.00	0.86	3
CP senior	1.00	1.00	1.00	2
Commercial	0.50	1.00	0.67	1
DP	1.00	1.00	1.00	2
DP senior	1.00	1.00	1.00	1
avg / total	0.70	0.82	0.75	11

```

Logistic Regression :
2018-01-29 13:24:31,680 - INFO - Training done in 0.000s
2018-01-29 13:24:31,680 - INFO - Testing Logistic Regression
2018-01-29 13:24:31,680 - INFO - Testing done in 0.000s
Score on training : 0.860465
Score on testing : 0.909091

```

	precision	recall	f1-score	support
Architecte	1.00	1.00	1.00	2
CP	1.00	1.00	1.00	3
CP senior	1.00	1.00	1.00	2
Commercial	1.00	1.00	1.00	1
DP	1.00	0.50	0.67	2
DP senior	0.50	1.00	0.67	1
avg / total	0.95	0.91	0.91	11

```

Knn Test accuracy score : 0.8181818181818182
Logistic Test accuracy score : 0.9090909090909091

```