Prédire la variété d'un iris Christophe Viroulaud

mettre iris.zip sur site

Prédire la variété d'un iris

Prédire la variété d'un iris

Christophe Viroulaud

Première NSI

Prédire la variété d'un iris

Il est possible d'utiliser ces données pour pouvoir classifier un iris inconnu. Comment produire déduction à partir de données existantes? \to prémisse IA : machine learning

En 1936, le biologiste *Ronald Fisher* a rassemblé les mesures de trois espèces d'iris.







Iris setosa

Iris versicolor

Iris virginica

Prédire la variété d'un iris

Problématique

Utiliser les donnée

informations

Prédire la variété

Algorithme kNN

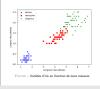
sentation nstruction de l'algorithm

truction de l'algorithme émentation

Prédire la variété d'un iris

Comment prédire une information nouvelle à partir de données brutes ?

Prédire la variété d'un iris Utiliser les données Présentation graphique des informations



- 1. Une représentation graphique des informations apporte une compréhension plus éclairante.
- 2. Il apparaît que les mesures d'un iris peuvent permettre de déterminer leur variété.

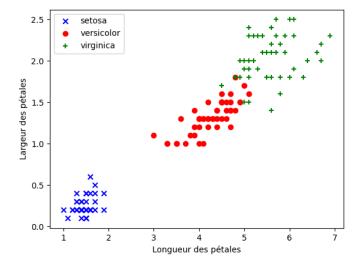


FIGURE – Variétés d'iris en fonction de leurs mesures

Prédire la variété d'un iris

Problématique

Utiliser les données Présentation graphique des

informations Prédire la variété

gorithme kNN

onstruction de l'algorithme

Activité 1 :

1 Détermine la variété des rés subparts :

Largeur 10.5 25 1.55 0.85

2 Proposer un entétode pour affectuer un choix dans les cas ambigus.

Utiliser les données pour prédire

Utiliser les données pour prédire

Activité 1 :

1. Déterminer la variété des iris suivants :

longueur	1	6	5.1	2.5
largeur	0.5	2.5	1.55	0.85

2. Proposer une méthode pour effectuer un choix dans les cas ambigus.

Prédire la variété d'un iris

Problématique

Otiliser les données

Prédire la variété

Algorithme kNN

résentation

variété

Correction

longueur	1	6	5.1	2.5
largeur	0.5	2.5	1.55	0.85

virginica

setosa

ambigu

ambigu

Prédire la variété d'un iris

Problématique

Utiliser les données

informations

Prédire la variété

gorithme Kiviv

truction de l'algorithme émentation Pour déterminer la variété d'un ints inconnu :

> regarder la variété d'un nombre k de voicins

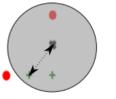
Méthode des k plus proches voisins

k Nearest Neighbors

Méthode des k plus proches voisins

Pour déterminer la variété d'un iris inconnu :

▶ regarder la variété d'un nombre k de voisins,



Prédire la variété d'un iris

Problématique

Utiliser les données

informations

Présentation

truction de l'algorithme émentation Méthode des k plus proches voisins

Pour détermer la variét d'un in incomu :

> regarder la variét d'un nombre à de voisins,

** régarder la variét d'un nombre à de voisins,

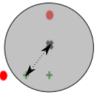
** Attribur à la finer incomus, la variété la plus présente pares ou à voisins.

k Nearest Neighbors

Méthode des k plus proches voisins

Pour déterminer la variété d'un iris inconnu :

regarder la variété d'un nombre k de voisins,



▶ attribuer à la fleur inconnue, la variété la plus présente parmi ses *k* voisins.

Prédire la variété d'un iris

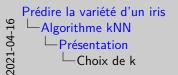
Problématique

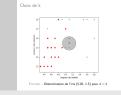
Utiliser les données

Informations

laorithma kNN

Présentation





Choix de k

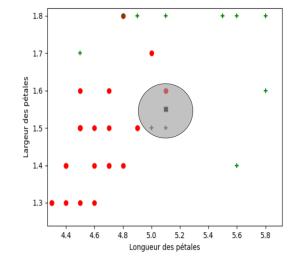


FIGURE – Détermination de l'iris (5.05, 1.5) pour k = 3

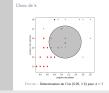
Prédire la variété d'un iris

Problématique

ésentation graphique of formations

Présentation

ruction de l'algorithme mentation 2021-04-16



- 1. avantages de kNN : simple d'implémentation et résultats avec bon taux de réussite quand on a un gros échantillon
- 2. inconvénients : sensible au bruit (données mal étiquetées), sensible à l'échelle de chaque dimension

Choix de k

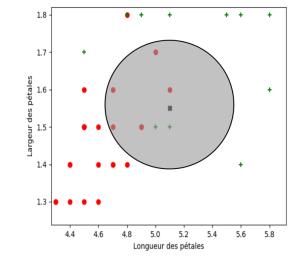


FIGURE – Détermination de l'iris (5.05, 1.5) pour k = 7

Prédire la variété d'un iris

Problématique

Utiliser les donné

Présentation graphique of informations

goritime kiviv

Présentation

émentation

Prédire la variété d'un iris -Algorithme kNN -Présentation

L'algorithme kNN est une méthode d'apprentissage suétiquetées sur lequel il va pouvoir s'entraîner et définir un

- 1. non supervisé : données pas étiquetées
- 2. par renforcement : par récompense

Complément

L'algorithme kNN est une méthode d'apprentissage supervisé : l'algorithme reçoit un ensemble de données déjà étiquetées sur lequel il va pouvoir s'entraîner et définir un modèle de prédiction.

Prédire la variété d'un iris

Présentation

10 / 24

Prédire la variété d'un iris Algorithme kNN Construction de l'algorithme Calcul de la distance



Calcul de la distance

Le plus naturel ici est de prendre la distance à vol d'oiseau ou plus formellement la distance euclidienne.

$$d = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

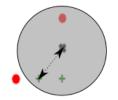


FIGURE – distance euclidienne

Prédire la variété d'un iris

Problématique

Itiliser les donnée

Présentation graphique de informations

Predire la variete

gorithme kNN

2021-04-16

Calcul de la distance

$$d = |x_A - x_B| + |y_A - y_B|$$

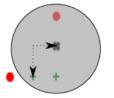


FIGURE – distance de Manhattan

Prédire la variété d'un iris

Problématique

iliser les données

rédire la variété

lgorithme kNN résentation

Activité 2 : Écrire en langage naturel, l'algorithme kNN.

Activité 2 : Écrire *en langage naturel*, l'algorithme kNN.

olématique

Prédire la variété d'un iris

Présentation graphique des informations Prédire la variété

> gorithme kNN sentation

- ▶ Charger les données dans le programme.
- Charger les données dans le programme.

 Choisir k.
- Stocker les mesures de la fleur inconnue.
- ► Calculer la distance euclidienne entre la fleur inconnue
- et tous les autres iris.

 Sélectionner les k plus proches iris (en distance) de la
- fleur inconnue.

 In Affortar la varieté maioritaire des le plus marches ide (es
- Affecter la variété majoritaire des k plus proches iris (en distance) à la fleur inconnue.

- ► Charger les données dans le programme.
- ► Choisir k.
- ► Stocker les mesures de la fleur inconnue.
- ► Calculer la distance euclidienne entre la fleur inconnue et tous les autres iris.
- ► Sélectionner les *k* plus proches iris (en distance) de la fleur inconnue.
- Affecter la variété majoritaire des *k* plus proches iris (en distance) à la fleur inconnue.

Prédire la variété d'un iris

Problématique

Présentation graphique des informations

lgorithme kNN

Construction de l'algorithme

nplémentation

Pour charger les données on utilisera la bibliothèque cav.

Pour charger les données on utilisera la bibliothèque *csv*.

Implémentation

Prédire la variété

d'un iris

Activité 3 :

- Télécharger le dossier compressé iris.zip sur le site https://cviroulaud.github.io
- Ouvrir le fichier data-iris.csv avec un tableur pour observer les données.
 Ouvrir le fichier iris-eleve.py.

Activité 3 :

- 1. Télécharger le dossier compressé *iris.zip* sur le site https://cviroulaud.github.io
- 2. Ouvrir le fichier *data-iris.csv* avec un tableur pour observer les données.
- 3. Ouvrir le fichier *iris-eleve.py*.

Prédire la variété d'un iris

Problématique

Utiliser les donnée

Prédire la variété

gorithme kNN

| patal_length | patal_width | species | 1.4 | 0.2 | setona | 1.4 | 0.2 | setona | 1.3 | 0.2 | setona |

Correction

Correction

3 attributs

petal_length	petal_width	species
1.4	0.2	setosa
1.4	0.2	setosa
1.3	0.2	setosa

Prédire la variété d'un iris

Problématique

Utiliser les données

Prédire la variété

gorithme kNN

sentation

Prédire la variété d'un iris -Algorithme kNN –Implémentation

- 4. Compléter la fonction charger_donnees en
 - utilisant les informations du fichier csv. Compléter la fonction distance qui calcule le carré de la distance euclidienne entre deux points du

 - Compléter enfin la fonction trouver variete. Le dictionnaire compteur_voisins compte le nombre d'apparitions de chaque variété parmi les à voisins.

Activité 3:

- 4. Compléter la fonction charger_donnees en utilisant les informations du fichier csv.
- 5. Compléter la fonction distance qui calcule le carré de la distance euclidienne entre deux points du plan.
- 6. Compléter la fonction *calculer distances*.
- 7. Compléter enfin la fonction trouver variete. Le dictionnaire compteur_voisins compte le nombre d'apparitions de chaque variété parmi les k voisins.

Prédire la variété d'un iris

```
Correction

| of department(our_ficher or) -> dec

ficher = question_ficher)
| dec_general_tour_ficher)
| dec_general_tour_ficher)
| dec_general_tour_ficher]
| dec_general_tour_ficher
```

```
def charger_donnees(nom_fichier: str) -> dict:
         fichier = open(nom fichier)
         data iris = csv.DictReader( fichier, delimiter = ",")
         dico varietes = {"setosa": [], "versicolor": [],
            virginica ": []}
        # Pour chaque ligne de données
        for iris in data iris:
            # Stocke la longueur et la largeur sous forme de
                tuple de flottants
            dico_varietes [ iris ["species"]]. append(
                (float (iris ["petal_length"]), float (iris ["
                    petal width"])))
         fichier . close ()
10
         return dico_varietes
11
```

Prédire la variété d'un iris

Problématique

Jtiliser les données

Prédire la variété

ritnme KIVIV ntation



[1])**2

def distance(connu: tuple, inconnu: tuple) -> float:

return (connu[0]—inconnu[0])**2+(connu[1]—inconnu



Prédire la variété d'un iris

```
I del calcular_distances (decreases dels_income top

statemen = []

for non, manuer in demonstratem();

for into in manuer

for into in manuer

dels dels non, income)

dels dels non, income dels dels non

return dels non

return dels none.
```

```
def calculer_distances (donnees: dict, inconnu: tuple) ->
        list:
        distances = []
        for nom, mesures in donnees.items():
            for iris in mesures:
 4
                d = distance( iris , inconnu)
                distances .append((nom, d))
 6
        # trie les iris en fonction de la distance
 8
        distances . sort (key=lambda fleur: fleur [1])
 9
        return distances
10
```

Problématique

Jtiliser les donnée

rédire la variété

orithme kNN

```
Correction

2 of source, price (it. int., distance, its ) -> int.

2 of source, price (it. int., distance, its ) -> int.

3 of source, interference of chapte switch

4 complete, source, interference of chapte switch

5 of source, interference, interference of chapte switch

6 of source, interference, interference of chapte switch

6 of source, interference, interference of chapter switch

10 of source, interference of source, interference of chapter switch

11 of source, interference of source, interference, interference of chapter switch

12 of source, interference of chapter switch

13 of source, interference of chapter switch

14 of source, interference of chapter switch

15 of source, interference of chapter switch

16 of source, interference of chapter switch

17 of source, interference of chapter switch

18 of source, interference o
```

```
def trouver_variete(k: int, distances: list) -> str:
        # compte le nombre d'occurences de chaque variété
        compteur\_voisins = \{\}
        for i in range(k):
 4
            nom = distances[i][0]
            if nom in compteur_voisins:
 6
                compteur\_voisins[nom] += 1
 8
            else:
                compteur\_voisins[nom] = 1
10
        # recherche la variété avec la plus grande valeur
11
           dans compteur voisins
        \max_{i} = 0
12
13
        nom maxi = 0
        for nom, quantite in compteur voisins.items():
14
            if quantite > maxi:
15
                maxi = quantite
16
                nom maxi = nom
17
18
        return nom maxi
```

Prédire la variété d'un iris

roblématique

illiser les données ésentation graphique des formations

rithme kNN

Présentation
Construction de l'algorithm
Implémentation

Prédire la variété d'un iris —Algorithme kNN —Implémentation

Activité 3 :

- Tester la fonction avec k = 3 puis k = 7, puis pour les autres iris de l'activité 1.
- Pour les plus avancés: Modifier le code pour tester un ensemble de 10 iris inconnus. De plus chaque iris déterminé sera ajouté au dictionnaire varietes afin d'augmenter l'apprentissage de l'algorithme.

Activité 3 :

- 8. Tester la fonction avec k = 3 puis k = 7, puis pour les autres iris de l'activité 1.
- 9. Pour les plus avancés : Modifier le code pour tester un ensemble de 10 iris inconnus. De plus chaque iris déterminé sera ajouté au dictionnaire varietes afin d'augmenter l'apprentissage de l'algorithme.

Prédire la variété d'un iris

Problématique

Utiliser les données

Prédire la variété

gorithme kNN

```
Correction

1 (x = 3)
2 close = (5.1.155)
3 close = charge_domens("data-inic.co")
4 varies = charge_domens("data-inic.co")
5 closes = closes_domens(varies close)
5 varies = traver_varies(h_dataron_close)
1 point("La varies(h_dataron_close))
2 point("La varies(h_dataron_close))
```

```
1  k = 3
2  cible = (5.1, 1.55)
3
4  varietes = charger_donnees("data-iris.csv")
5  distances_cible = calculer_distances ( varietes , cible )
6  variete = trouver_variete(k, distances_cible )
7
8  print("La variété est ", variete)
```

Prédire la variété d'un iris

Problématique

Jtiliser les données

informations

lgorithme kNN

entation

```
k = 3
    cibles = [(1,0.5),(6,2.5),(5.1, 1.55),(2.5,0.85),(3,2),
               (6,1.2), (2.1,1), (3.2,1.5), (3.5,2.5), (4,1)
 3
 4
    varietes = charger_donnees("data-iris.csv")
    for iris_inconnu in cibles:
        # trouve la variété
         distances cible = calculer distances (varietes,
           iris inconnu)
         variete = trouver_variete(k, distances_cible)
        print (f"La variété de {iris_inconnu} est {variete}.")
        # ajout de cible au dictionnaire des données
11
         varietes [ variete ]. append(iris_inconnu)
12
```

Prédire la variété d'un iris

Problématique

Jtiliser les données

Prédire la variété

orithme kNN