LABORATORIUM 8

Wszystkie użyte biblioteki i funkcje

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

2.1 Implementacja klasyfikatora k-NN

1Zaimplementuj algorytm k-NN jako klasę w języku Python.

\*Konstruktor

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Konstruktor przyjmuje liczbę sąsiadów oraz wartość bool która definiuje czy będziemy korzystali z KD-drzew.

\*Funkcja fit

Obraz zawierający tekst, Czcionka, pismo odręczne, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Jest to metoda do uczenia modelu zapisuje dane wejściowe, cechy oraz atrybuty decyzyjne.

\*Metoda do dokonywania predykcji

-Część bez KD-drzew

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Na początku tworzymy macierz o liczbie wierszy naszych próbek w zbiorze. Macierz posiada dwie kolumny. Pierwsza kolumna służy do przechowywania zmiennej zależnej (y\_train[i]), a druga do przechowania odległości danego punktu treningowego do punktu testowego X. Tworzymy pętle i inicjalizujemy macierz analogicznie dla każdego punktu testowego..

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

W Kolejnym etapie jest sortujemy macierz względem odległości euklidesowej, Dzięki temu na początku macierzy znajdują się te punkty treningowe które są najbliżej punktu testowego.

Z macierzy wybieramy tylko najbliższych sąsiadów (ich ilość przekazujemy na początku algorytmu w parametrze n\_neighbours). Następnie wyodrębniamy ich atrybuty klasowe.

Za pomocą funkcji np.unique zostaje zwrócona wartość unikalnych atrybutów oraz zostaje zliczona ich ilość wystąpień. Na końcu sortuje te atrybuty według częstośći występowania. (od największej do najmniejszej).Zwracamy najczęściej występujący atrybut wśród sąsiadów.

-Część z KD-Drzewem

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rozszerzamy implementacja, aby mogła korzystać z kD-drzew. By to wykonać przekształcamy nasze dane treningowe na tablice dwuwymiarową i umieszczamy je w strukturze KDTree z biblioteki sklearn.neighbours. Następnie korzystając z metody kdtree.query pobieramy indeksy najbliższych sąsiadów dla danego punktu. Dalsze etapy działania przebiegają analogicznie jak przy algorytmie bez KD-drzew.

Struktura KD-Drzew umożliwia nam sprawniejsze i bardziej wydajne znajdowanie najbliższych sąsiadów. Dzięki strukturze drzewiastej nie musimy przeszukiwać wszystkich punktów. Dla przykładu jeśli najbliższy możliwy punkt w jakiejś gałęzi jest dalej niż najdalszy ze znalezionego jakiegoś sąsiada. Algorytm ignoruje całą gałąź i szuka w gałęzi znalezionego sąsiada.

\*Metoda zwracająca wskaźnik jakości dopasowania:

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Metoda prosta w implementacji i zrozumieniu. Przewiduje atrybuty dla danych testowych jeśli atrybuty się zgadzają z oryginalnymi zwiększamy licznik hit o jeden. Następnie dzielimy przez ilość punktów testowych i przekształcamy na procenty.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.CAŁA KLASA

2.2 Implementacja k-NN dla regresji

CAŁA KLASA

Obraz zawierający tekst, zrzut ekranu, Czcionka, oprogramowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Różnica w algorytmie k-NN dla regresji polega na tym, że przewidujemy wartość liczbową (zmienną ciągłą), a nie klasę (atrybut) jak w przypadku klasyfikacji. Klasa knn\_regression dziedziczy po klasie dla klasyfikacji konstruktor i metode fit. W funkcji predict nastąpiło kilka zmian. Obliczamy średnia z wartości liczbowych naszych najbliższych sąsiadów co jest naszym wynikiem końcowym.

Metoda score oblicza błąd średniokwadratowy. Obliczamy różnicę między oryginalnym a przewidzianym punktem przy czym podnosimy do potęgi, żeby zniwelować minusa. Obliczamy średnią wszystkich różnic i pierwiastkujemy.

3.KLASYFIKACJA

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.Zadanie 1. Wygeneruj dane uczące przy pomocy metody sklearn.datasets.- make classification.

Zadanie 2. Dokonaj klasyfikacji przy pomocy metody k-nn.



Zadanie 3. . Zwizualizuj dane oraz granicę separacji w przestrzeni 2D. Granica separacji może zostać zobrazowana przy pomocy metody contour. W celu jej narysowania należy wygenerować regularną siatkę punktów (metoda meshgrid), a następnie dla każdego węzła siatki obliczyć odpowiedź algorytmu k-nn. P

Obraz zawierający tekst, zrzut ekranu, Czcionka, oprogramowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Generujemy siatkę punktów za pomocą funkcji np.meshgird, przekazując jej tablice reprezentujące wartości cech w zadanym kroku h. Kolejnym krokiem jest stworzenie macierzy z przewidzianymi zmiennymi zależnymi dla każdego punktu siatki. Wizualizujemy punkty za pomocą plt.scatter i ich granice zmian atrybutów funkcją plt.contour

REZULTAT

Obraz zawierający mapa, diagram, zrzut ekranu, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zadanie 4. Wczytaj dane iris oraz rozdziel je na cześć wejściową (X) i decyzje (Y ), a następnie dokonaj ich klasyfikacji.

Obraz zawierający tekst, Czcionka, zrzut ekranu, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zadanie 5.Zwizualizuj dane przy pomocy metody PCA rzutując dane na dwie pierwsze składowe główne. Umieść na wykresie granicę separacji.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Rzutujemy dane na dwie pierwsze składowe za pomocą metody PCA ,redukując wymiarowość z 4D do 2D. Następnie, analogicznie do zadania nr 3, generujemy siatkę punktów dla przestrzenie 2D. Przewidujemy ich atrybuty z dodatkowym uwzględnieniem przekonwertowania punktów siatki do oryginalnej przestrzeni 4D przed użyciem metody predict. Na końcu wizualizujemy wykres 2D.

REZULTAT

Obraz zawierający tekst, diagram, mapa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zadanie 6. Przy pomocy kroswalidacja leave-one-out przetestuj jak algorytm zachowuje się przy różnych wartościach parametru k. Wyniki wypisz na ekran w postaci tabeli.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Testujemy algorytm na bazie iris. Wybrałem 9 różnych parametrów k różniące się od siebie. Dla każdego parametru dokonuje klasyfikacji dla każdej próbki występujących w danych treningowych. Usuwam próbkę z danych treningowych jak i atrybutów za pomocą funkcji np.delete . Następnie przewiduje jej atrybut. Podobnie jak przy metodzie score dla klasyfikacji obliczam procentową poprawność algorytmu dla przewidzianej każdej próbki.

REZULTAT

Obraz zawierający tekst, Czcionka, zrzut ekranu, czarne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Z rezultatu można wywnioskować, że liczba sąsiadów nie może być zbyt mała (model może podejmować decyzje na podstawie przypadku, zbyt czuły na szumy) jak i nie może być nader duża (Model nie zauważy lokalnych różnic).

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.Zadanie 7. Porównaj czas działania algorytmu w wersji podstawowej i wersji korzystającej z kD-Drzew.

Dzielimy dane i zmienne zależne na treningowe i testowe za pomocą train\_test\_split (W skali 60%, 40%) Testujemy dla trzech różnych sąsiadów i mierzymy czas wykonania algorytmu. Dodatkowo wyświetlane zostaną wyniki dokładności w celu przetestowania metody.

REZULTAT

Obraz zawierający tekst, Czcionka, zrzut ekranu, czarne i białe

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Można zauważyć, że użycie KD-drzew przyśpiesza działanie algorytmu w niektórych przypadkach nawet dwukrotnie. Metoda score dla obu przypadków wyświetla identyczny wynik.

4.REGRESJA

Zadanie 1. Wygeneruj dane uczące (2D) przy pomocy metody sklearn.datasets.- make regression.

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zadanie 2. Dokonaj regresji przy pomocy metody k-nn.

Zadanie 3. Zwizualizuj dane oraz odpowiedź modelu (linia trendu). Wypisz na ekran błąd popełniany przez model.

Obraz zawierający tekst, zrzut ekranu, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Na początku podzieliłem dane na treningowe i testowe by obliczyć metodą score średni błąd.

Obraz zawierający tekst, zrzut ekranu, Czcionka, oprogramowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Później korzystając z siatki zwizualizowałem linie trendu. Oraz umieściłem nasze dane na wykresie.

REZULTAT

Obraz zawierający tekst, zrzut ekranu, diagram, mapa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Zadanie 4. Wczytaj dane boston (sklearn.datasets.load boston).



Niestety z powodu nowszej wersji biblioteki sklearn nie mogę wczytać danych boston. Od wersji 1.2 biblioteka boston została usunięta z powodów etnicznych. W zamian zaimportowałem bazę podobną do bouston. Z powodu wielkości bazy i problemów z załadowaniem danych pobrałem tylko pierwsze 200 elementó

Zadanie 5. Przy pomocy 10-krotnej krzyżowej walidacji przetestuj model dla różnych wartości parametru k. Wyniki wypisz na ekran w postaci tabeli

Obraz zawierający tekst, zrzut ekranu, Czcionka, oprogramowanie

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Podzieliłem dane na 10 równych podzbiorów za pomocą funkcji KFold. Następnie wyliczyłem średni błąd dla różnej ilości sąsiadów i wyświetliłem w tabeli.

REZULTAT

Obraz zawierający tekst, Czcionka, zrzut ekranu, czarne

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.