Phép đánh giá P ( performance P)

Để kiểm tra hiệu năng cảu một thuật toán ML . chúng ta cần phải thiết kế các phép dánh giá có thể đo dặc được

Kinh nghiệm E ( experience E)

Việc huấn luyện các mô hình ML có thể coi là việc cho chúng trải nghiệm trên các tập dữ liệu – chính là training set

Các tập dữ liệu khác nhau sẽ choc ac smoo hình các trải nghiệm khác nhai

Chất lượng của các tập dữu liệu này cũng ảnh hưởng tới hiu năng của mô hình hính

Thị giác máy tính (computer vision) là một lingx vực đã và đang phát triễn mạnh mẽ trên thế giới hiện nay.Khái niệm về thị giác máy tính -Computer vision có liên quan tới nhiều ngành học và có nhiều hướng nghiên cứu khác nhau

Kể từ những năm 70 của thế kỷ 20 khi khả năng tính toán của máy tính trở nên mạnh mẽ . NÓ có thể giải quyết được những công việc khó như xử lý các tập ảnh , các đoạn video….

Thị giác máy tính được nghiên cứu và phát triển cho đến ngày nay

Chương III : K-NEAREST NEIGHBOR cho phân lớp ảnh

Thuật toán K-NEAREST knn là một kỹ thuật học máy phân loại học có giám sát phổ biến được sử dụng cho các nhiệm vụ phân loại và hồi quy

KNN dựa trên ý tưởng rằng các điểm dữ liệu tương tự có xu hướng có nhãn hoặc giá trị tương tự

Trong giai đoạn huẩn luyện , thuật toán KNN lưu trữ toàn bộ tập dữ liệu huấn luyện làm tài liệu tham khảo

Khi đưa ra dự đoán , nó sẽ tính toán khoảng cách giữa điểm dữ liệu đầu vào cà tất cả các mẫu huấn luyện bằng cach sử dụng thước đo khoảng cách đã chọn chẳng hạn như khoảng cách Euclide

KNN xác định K hàng xóm gần nhất với điểm dữ liệu đầu vào dựa trên khoảng cách của chúng

Đối với hồi quy nó tính toán giá trị trung binh hoặc trung binh có trọng số của các giá trị mục tiêu của K lân cần để dự đoán giá trị cho điểm dữ liệu đầu vào

Với KNN trong bài toán classification , label của một điểm dữ liệu mới (hay kết quả của câu hỏi trong bài thi ) được suy ra trực tiếp từ K điểm dữ liệu gần nhất trong training test

Nhãn lớp của 1 test data có thể được quyết định bẳng bỏ phiếu đa số ( major voting) giữa các điểm gần nhất , hoặc nó có thể được suy ra bằng cách đánh trọng số khác nhau cho mỗi trong các điểm gần nhất đó rồi suy ra nhãn lớp.

Sự khác biệt giữa các thuật ngữ này là về mặt kỹ thuật , “ bỏ phiếu đa số “ yêu cầu dda số lớn hơn 50% điều này chủ yếu có hiệu quả khi chỉ có hai loại

Khi có nhiều lớp

VD: ba loại ko nhất thiết cần 50% phiếu bầu để đưa ra kết luận về một lớp: có thể chỉ định nhãn lớp với số phiếu bầu lớn nhãn và lớn hơn 33,33%.

K-nearest neighbor là một trong những thuật toán supervised-learning đơn giản nhất ( mà hiệu quả trong một vài trường hợp trong machine learning)

Khi training thuật toán này không học một điều gì từ dữ liệu training ( đầy cũng là lú do thuật toán này được xếp vào loại lazy learning), mọi tính toán được thực hiện khi nó cần dự đoán kết quả của dữ liệu mới.

Phân lớp ảnh : KNN có thể được sử dụng để phân loại hình ảnh vào các lớp khác nhau , chẳng hạn như phần lớp ảnh khuôn mặt , phân loại loài cây dữa trên hình ảnh lá , hoặc phân loại sản phầm .

Nghiên cứu hành vi khách hàng : KNN có thể được sử dụng để xác định các khách hàng có hành vi tương tự VD: nếu biết khách hàng A mưa các sản phẩm X,Y và Z , có thể sự dụng KNN để tìm cac khách hàng khác có hành vi tương tự và gợi ý sản phầm

Phát hiện bất thường : KNN có thể được sử dụng để phát hiện bất thương trong dữ liệu . Nếu một diểm dữ liệu có khoảng cách lớn đến K điểm gần nhất , nó có thể được xem là một bất thường .

Gợi ý sản phẩm : KNN có thể được sử dụng để gợi ý sản phẩm cho người dùng dựa trên hành vi mua sắm hoặc xem sản phẩm trước đó

Dự đoán giá trị :KNN có thể dưỡc sử dụng để đoán giá trị của một biến dựa trên giá trị của các biến khác ,VD , dự đoaán giá nhà dựa trên diện tích , số phòng và vị trí

Phân tích dữ liệu y tế : Trong lĩnh vực y tế ,KNN có thể dược sử dung đê phân loại các bệnh , phát hiện dấu hiệu bất thường trong hồ sơ bệnh án , hoặc gợi ý diều trị dựa trên tiền sử bệnh của bênh nhân

Ứng dung tronhg lĩnh vựa xử lý ảnh và thị giác máy tính : KNN có thể đưọc sử dụng trong việc nhận diện khuôn mặt , phát hiện đối tượng trong ảnh , hoặc phân lớp các ký tự trong các hệ thống OCR(optical character regconition)

Ứng dụng trong lĩnh vực giọng nói :KNN có thể được sử dụng để nhận diện giọng nói và phân biệt giữa các loại giọng nói khác nhau

Phân loại văn bản : KNN có thể được sử dụng để phân loại văn bvanr vào các danh mục khác nhau , chẳng hạn như phân loại email là spam hoặc không spam

Phân lớp chữ viết tay

Phân loại đối tượng trong hình ảnh

Nhận diện biển báo giao thông

Phân loại dấu hiệu bệnh trên ảnh y khoa

Phân loại loài vật từ hình ảnh

Phân lớp sản phẩm trên của hàng trực tuyến

**Ưu điểm của KNN**

Dễ triễn khai và dễ hiểu

Không yêu cầu giả định về phân phoois

Hiệu quả với dữ liệu lớn

Có khả năng cập nhật dữ liệu

Khả năng hoạt động với đa dạng các loại đặc trưng

**Nhược điểm KNN**

Nhạy cảm với nhiễu

Yêu cầu lựa chọn giá trị K phù hợp

Tính toán đọ phức tạp cao: KNN phải tinh toán khoảng cách giữa điểm dữ liệu và tất cả các điểm dữ liệu khác trong tập huấn luyện .Điều này có thể làm cho thuật toán trở nên tốn thời gian đối với tập dữ liệu lớn

Không có khả năng học được cấu trúc dữ liệu ẩn :KNN không có khả năng học đưỡ cấu trúc dữ liệu ẩn trong dữ liệu .Nó chỉ dựa vào thông tin từ các điểm gần nhất

Khó xử lý với dữ liệu không cân bằng :KNN ko xử lý tốt với dữ liệu ko cần bằng , nơi một lớp có số lượng lớn hon hẳn so với các lớp khác.

Sử dụng ngôn ngữ python ứng dụng KNN để giải quyết 1 bài toán

In put sensing device interpreting device out put

conda - -version ( xem phien ban )

conda update conda (

conda create - -name …… biopython ……… (tao ten moi truong)

conda activate ………. ( doi ten moi truong can toi )

Conda info - -envs (xem danh sacsh cac moi truong)

>cd ( chuyeen doi o luu)

Conda create - - name snakes python=3.9 ( taoj moi truong cung vs phien ban python)

Python - - version (kiem tra python)

Conda search beautifulsoup4 (kiem tra package)

Conda install tensorflow ( cài dat package)

Conda list ( kiêm tra package)

Conda list -n tenmoitruong (kiêm tra pakage có trong 1 deactivated environment )

Conda install - -name CV scipy ( cài scipy vào moi trương CV)

Conda install scipy ( cài package vào moi truong hien tai)

Conda install scipy=0.15.0 (cập nhập package phien ban mongmuon)

Conda install scipy numpy ( cài đặt nhiều package)

Conda install scipy=0.15.0 curl=7.26.0 (cài version cụ thể cho nhiều package cùng 1 lúc )

Pip install tensorflow (

Conda update numpy ( update 1 package)

Conda update python (update python)

Conda update conda (update conda)

Nghe 070

Bài thầy dũng crawl dữ liệu thứ 3

Thực tập 27/11

Bài toán regression xem người đó có đủ người điều kiện vay hay ko thứ 6

Cài đặt các thư viện thứ 2

Cơ sở dữ liệu phân tán ( SQL )