

AI & Applications

Lecturer: Hoang Duc Quy

Email: quyhd@ut.edu.vn

Department: Institute of Information Technology and
Electrical- Electronic Engineering (IT3E)

Outline

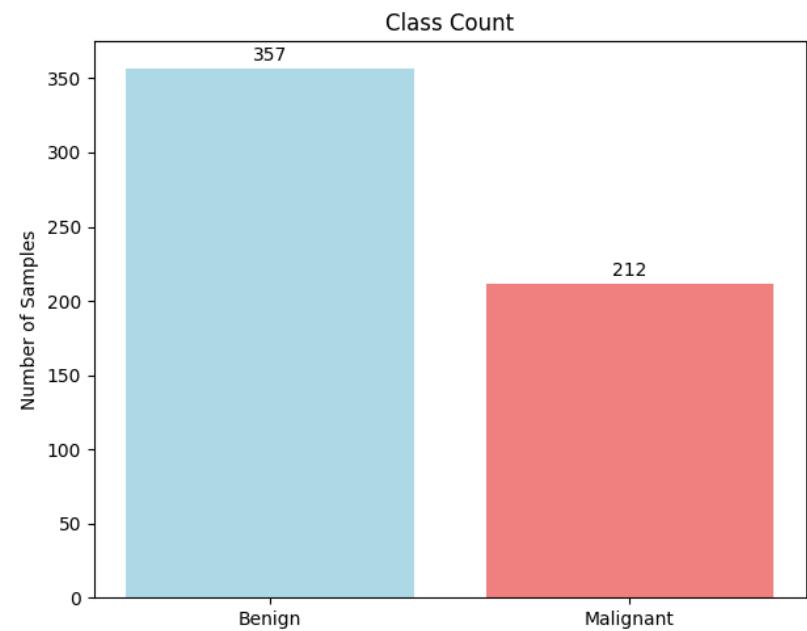
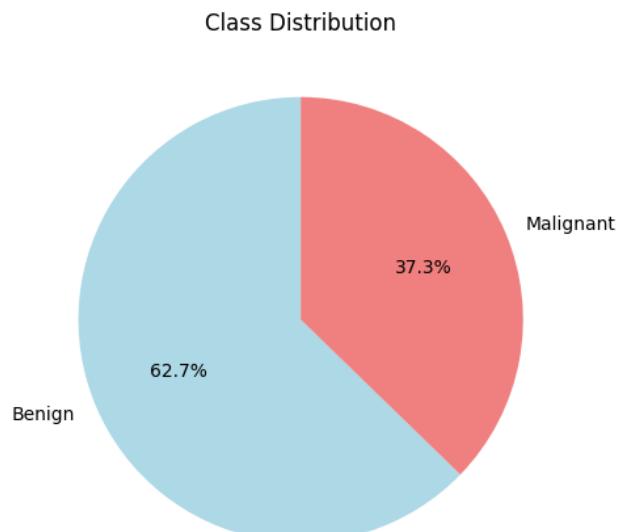
- Project 1: Áp dụng thuật toán KNN trong việc dự đoán ung thư
- Project 2: Áp dụng Linear Regression trong việc dự đoán giá nhà ở California
- Project 3: Tùy chọn

Project 1

Áp dụng thuật toán KNN trong việc dự đoán ung thư

introduction

- Tập dữ liệu này chứa thông tin về các đặc trưng liên quan đến khối u ung thư vú, cùng với chẩn đoán của chúng
- Số lượng đặc trưng: 30
- Mục tiêu (nhãn):
 - 0: 212 mẫu (ác tính)
 - 1: 357 mẫu (lành t



Goal

- 1) Xây dựng mô hình KNN để phân loại mẫu đầu vào là LÀNH TÍNH hay AC TÍNH
- 2) Kiểm định các trường hợp hệ số k khác nhau, xác định hệ số k tốt nhất
- 3) Tiêu chí đánh giá là độ chính xác được xác định bằng công thức sau:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Trong đó, TP, TN, FP, FN lần lượt là các trường hợp dương tính thật, âm tính thật, dương tính giả và âm tính giả.

LƯU Ý: Ở tiêu chí 3, giữa các nhóm cùng chung một đề tài, nhóm nào có chỉ số acc cao nhất sẽ có điểm + (có minh chứng)

Material

- Các nhóm sẽ được cung cấp một file source code hỗ trợ việc load và phân chia dữ liệu cơ bản

TP – True positive

- Dương tính thật: một trường hợp được coi là dương tính thật khi mô hình dự đoán dương tính và nhãn thực tế là dương tính.
- Trong trường hợp ở Project này, lớp dương tính là 0 (ác tính)

TN – Âm tính thật

- Âm tính thật: một trường hợp được coi là âm tính thật khi mô hình dự đoán âm tính và nhãn thực tế là âm tính.
- Trong trường hợp ở Project này, lớp âm tính là 1 (lành tính)

Fp – Dương tính giả

- Dương tính giả: một trường hợp được coi là dương tính giả khi mô hình dự đoán dương tính nhưng nhãn thực tế là âm tính.
- Trong trường hợp ở Project này, lớp dương tính là 0, âm tính là 1

FN – âm tính giả

- Âm tính giả: một trường hợp được coi là âm tính giả khi mô hình dự đoán âm tính nhưng nhãn thực tế là dương tính.
- Trong trường hợp ở Project này, lớp dương tính là 0, âm tính là 1

Project 2

Áp dụng Linear Regression trong việc dự đoán giá nhà ở California

introduction

- Dự án này nhằm mục tiêu dự đoán giá nhà tại bang California dựa trên một số đặc điểm như vị trí địa lý, mật độ dân cư, thu nhập trung bình, số phòng, số hộ dân...
- Số lượng đặc trưng: 8
- Mục tiêu (nhᾶn): giá nhà (PRICE)

Project Goal

Apply LR to predict house prices in California

Dataset (example)

- **MedInc:** Median income in block group (in tens of thousands of dollars)
- **HouseAge:** Median house age in block group (in years)
- **AveRooms:** Average number of rooms per household
- **AveBedrms:** Average number of bedrooms per household
- **Population:** Block group population
- **AveOccup:** Average number of household members
- **Latitude:** Block group latitude
- **Longitude:** Block group longitude
- **PRICE:** Median house value (in hundreds of thousands of dollars)

```
MedInc,HouseAge,AveRooms,AveBedrms,Population,AveOccup,Latitude,Longitude,PRICE  
8.3252,41.0,6.984126984126984,1.0238095238095237,322.0,2.5555555555555554,37.88,-122.23,4.526  
8.3014,21.0,6.238137082601054,0.9718804920913884,2401.0,2.109841827768014,37.86,-122.22,3.585  
7.2574,52.0,8.288135593220339,1.073446327683616,496.0,2.8022598870056497,37.85,-122.24,3.521  
5.6431,52.0,5.8173515981735155,1.0730593607305936,558.0,2.547945205479452,37.85,-122.25,3.413  
3.8462,52.0,6.281853281853282,1.0810810810810811,565.0,2.1814671814671813,37.85,-122.25,3.422
```

Task

- 1) Xây dựng mô hình Linear regression (LR) với đầu vào nhiều đặc trưng
- 2) Kiểm định sự khác biệt giữa khi sử dụng mô hình thủ công và khi sử dụng mô hình có sẵn
- 3) Tiêu chí đánh giá là độ chính xác được xác định bằng công thức sau:

$$RMSE = \sqrt{1/2 (y - \hat{y})^2}$$

LƯU Ý: Ở tiêu chí 3, giữa các nhóm cùng chung một đề tài, nhóm nào có chỉ số RMSE cao nhất sẽ có điểm + (có minh chứng)

Support material

- You are provided some helper functions in **util.py**
- Dataset initialization is **DONE** in jupyter notebook file.

Project 3

Tùy chọn