Mỗi nhôm 4 người.2

Deadline: 25- 11

Submission format: power point trình bầy, mạng xây dựng thế nào, các parameters setup thế nào. Kết quả cuối cùng là thế nào. Nộp trên classroom

Supports:

* Question on classroom
* Vào chiều thứ 5 hàng tuần từ 15h đến 16h.

Hãy xây dựng một mạng Neuron để dự đoán một người có bị bệnh tim hay không. Nguồn dữ liệu lấy từ 5 nguồn khác nhau:

* Cleveland: 303 observations
* Hungarian: 294 observations
* Switzerland: 123 observations
* Long Beach VA: 200 observations
* Stalog (Heart) Data Set: 270 observations

Total: 1190 observations  
Duplicated: 272 observations

Final dataset: 918 observation

Để download và biết thêm thông tin thêm về Dataset:

<https://www.kaggle.com/fedesoriano/heart-failure-prediction>

Một số gợi ý về cách xây dựng và huấn luyện mạng:

* Để đơn giản các hàm activation mà mình sẽ sử dụng sẽ là: Relu cho các layer ở lớp giữa và Sigmoid dùng cho layer cuối cùng(output)
* Sử dụng các hàm có trong file dnn\_app\_utils\_v3.py như: linear\_forward, linear\_activation\_forward, compute\_cost, linear\_backward, linear\_activation\_backward, L\_model\_backward, predict. (Ví dụ cách dùng kết hợp các hàm này có thể xem hàm L\_layer\_model đã cung cấp)
* Dùng Panda để load dữ liệu từ file CSV. Trong DataSet ban đầu có 1 số cột mà giá trị là text, cần chuyển nó sang số vì dụ như cột Gender, ChesPainType,…(có thể dùng hàm get\_dummies của Panda)
* Sử dụng train\_test\_split của sklearn để tạo ra tập train và test riêng biệt
* Thay đổi số layers của mạng thông qua biến layers\_dims để tối ưu accuracy cao nhất(Có thể bắt đầu thử số layer là 3 và tăng dần)
* Thay đổi learning\_rate, num\_iterations để tối ưu accuracy cao nhất
* Hãy thử chuẩn hóa dữ liệu đầu vào để tối ưu accuracy cao nhất(ví dụ các input phải trong khoảng 0,1). Sau cùng thử sử dụng MinMaxScaler của sklearn để đánh giá accuracy.

Tạo báo cáo kết quả có accuracy cao nhất mà nhóm đã từng thử qua các params như: số layer và node trên mỗi layer, số lần iteration, learning\_rate, phương pháp chuẩn hóa dự liệu đầu vào

Note: Với setup cơ bản, có thể tạo ra một mạng có tỷ lệ train là 0.98% và test là 0.87% như hình phía dưới.

Graphical user interface, chart, line chart

Description automatically generated