

ĐỒ ÁN 2 - HỒI QUY LOGISTIC

1 Tập dữ liệu

1.1 Giới thiệu

- Đây là tập dữ liệu về một số loại ký hiệu xuất hiện trong ngôn ngữ ký hiệu.
- Tập dữ liệu được chia làm 2 phần: training set và test set.
- Training set gồm có 27000 mẫu. Test set có 7000 mẫu.
- Mỗi mẫu gồm có:
 - + Một bức ảnh trắng đen (grayscale), có kích thước 28×28 điểm ảnh (pixel). Chính giữa bức ảnh sẽ có hình ảnh về một ký hiệu.
 - + Nhân của bức ảnh sẽ có giá trị từ 0 đến 24, mỗi giá trị đó sẽ tương ứng với một loại ký hiệu nào đó.
- Đường dẫn để tải tập dữ liệu training set:
<https://drive.google.com/drive/folders/1lQz5986WF7a5W5IcLW8oekbrNGSbivire?usp=sharing>
- Cấu trúc của các tệp tin chứa tập dữ liệu:
 - + Gồm một file csv.
 - + Trong các file csv, mỗi mẫu sẽ gồm có các trường: label (nhãn) và các pixel (điểm ảnh).
 - + Điểm ảnh ở vị trí dòng i , cột j với $0 \leq i, j < 28$ sẽ tương ứng với pixel thứ $28*i + j + 1$.

1.2 Yêu cầu

Hoàn thành các yêu cầu sau theo thứ tự từ trên xuống dưới:

- Đọc được toàn bộ tập dữ liệu.
- Hiển thị hình ảnh và nhãn của bất kỳ mẫu nào trong tập dữ liệu. Hình ảnh cần được hiển thị như một bức ảnh grayscale, kích thước 28×28 pixel.

2 Hồi quy logistic

2.1 Yêu cầu

Hoàn thành các yêu cầu sau theo thứ tự từ trên xuống dưới:

- Xét một mẫu bất kỳ, trình bày cách chuyển dữ liệu về bức ảnh trong mẫu đó thành vector các số thực.
- Trình bày cấu trúc mô hình logistic: Đầu tiên, ta chuyển ảnh chuyển vector các số thực. Sau đó, ta thiết kế mô hình logistic như thế nào để áp dụng lên vector các số thực này?
- Sau khi đã thiết kế mô hình logistic, cho trước các tham số và đầu vào, trình bày quá trình dự đoán kết quả đầu ra.
- Sau khi đã huấn luyện các tham số xong, hiển thị độ chính xác của mô hình đối với một tập dữ liệu cho trước, trong đó, độ chính xác được đo bằng số mẫu dự đoán đúng chia cho tổng số mẫu (kích thước của tập dữ liệu).
- Lập bảng độ chính xác của mô hình đối với hai tập dữ liệu là training set và test set.