

Exercise 9 - Multilayer perceptron

Đặng Linh Anh

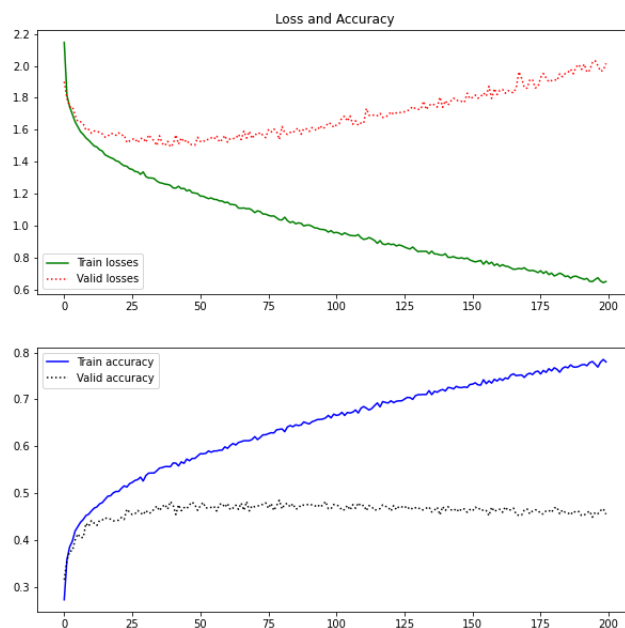
January 2021

1. Các bạn đọc hiểu các file code mẫu dùng MLP cho Fashion-MNIST và Cifar10
 - (a) 1.MLP-Fashion-MNIST-KerasData-Sigmoid.ipynb
 - (b) 2.MLP-Fashion-MNIST-KerasData-ReLU.ipynb
 - (c) 1.MLP-Cifar10-KerasData-Sigmoid.ipynb
 - (d) 2.MLP-Cifar10-KerasData-ReLU.ipynb
2. (Optional) Các bạn thực nghiệm với hàm tanh, và so sánh với model khi dùng Sigmoid

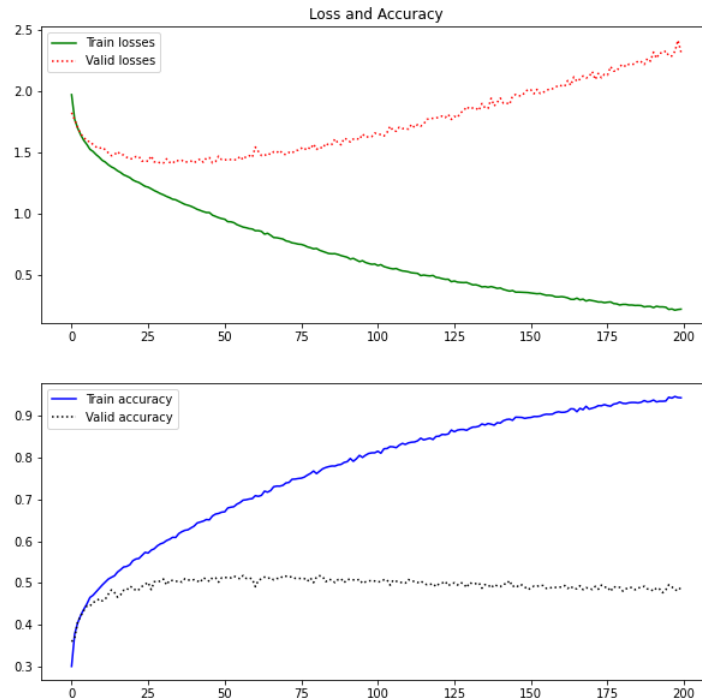
Lời giải

[LINK NOTEBOOK](#)

Kết quả khi huấn luyện model với hàm tanh:



Hình 1: Huấn luyện model sử dụng hàm **tanh** trên dataset Cifar-10



Hình 2: Huấn luyện model sử dụng hàm **sigmoid** trên dataset Cifar-10

Accuracy cho từng model đối với tập kiểm tra như sau:

- **tanh**: 45.85%
- **sigmoid**: 49.18%

Ta thấy, khi sử dụng hàm **sigmoid** cho activation, accuracy của model được cải thiện 3.33%. Tuy nhiên, trong cả 2 trường hợp, model đều bị overfitting sau khoảng 20 epochs (không phụ thuộc nhiều vào hàm initialiazer).

3. (Optional) Các bạn chọn 1 một model cố định, sau đó chạy thực nghiệm với các loại data normalization khác nhau. Sau đó các bạn so sánh kết quả và đưa ra các kết luận các bạn rút ra được.

Lời giải

Kết quả accuracy đạt được của từng model:

- **Gauss normalization**: 86.80%
- **Scale to $[0, 1]$** : 84.61%
- **Scale to $[-1, 1]$** : 86.15%

Khi dùng các normalization đưa data về distribution có dạng đối xứng qua điểm 0 như Gauss normalization và $[-1, 1]$, model đem lại kết quả tốt hơn so với khi scale về $[0, 1]$.

Khi xét Gauss normalization và $[-1, 1]$, Gauss đem lại kết quả tốt hơn vì model đang dùng Xavier initialization (là cách khởi tạo mặc định của keras) nên phân phối chuẩn của

data được học tốt hơn.

4. (Optional) Các bạn làm tương tự câu (4) cho 3 loại Weight Initialization đã học (Gaussian với mean=0, sdv=0.01; Glorot và He). So sánh kết quả khi thực nghiệm với cùng điều kiện.

Lời giải

Kết quả accuracy đạt được cho từng model:

- **Gauss initialization:** 86.25%
- **Glorot normal initialization:** 87.29%
- **He normal initialization:** 86.83%

Vì data được normalize theo Gauss normalization nên feed rất tốt cho cả 3 cách khởi tạo tham số trên. Đặc biệt đối với khởi tạo Glorot normal (Xavier), accuracy được đẩy lên cao nhất, với nguyên lý giữ cho $var(a^{(l-1)}) = var(a^{(l)})$ nên distribution ban đầu của data được giữ rất tốt xuyên suốt mạng neural nên mô hình học được tốt hơn.

[LINK NOTEBOOK CHO BÀI 3 VÀ 4](#)