

Lab Report

Experiment: Thí nghiệm tìm kiếm số lượng mẫu phù hợp

Title: Tìm kiếm số lượng mẫu tối ưu cho mô hình

Date: 02/12/2024 Name: Nguyễn Bá Sĩ Trâm

1. Mục tiêu thí nghiệm

- Đánh giá hiệu suất của mạng MLP (Multi-Layer Perceptron) với những tham số đã cho.
- Phân tích tác động của số lượng mẫu huấn luyện đến hiệu suất dựa trên chỉ số lỗi MSE (Mean Squared Error).

2. Cấu hình thí nghiệm

1. Dữ liệu:

- Tập dữ liệu huấn luyện được chia thành nhiều kích thước từ **200 đến 4000** mẫu, với bước tăng là 200.
- Tập xác thực (validation) và kiểm tra (test) được giữ nguyên.
- Dữ liệu gồm hai biến đầu vào (x_1, x_2) và một biến đầu ra (y).

2. Mô hình:

- MLP với 1 lớp ẩn (20 node ẩn), hàm kích hoạt ReLU.
- Hàm kích hoạt ở lớp đầu ra là **linear**.
- Hàm mất mát: **MSE**.
- Bộ tối ưu: SGD với learning rate: 0.01.

3. Callback:

- **EarlyStopping**: Dừng sớm khi validation loss không cải thiện sau 10 epoch.
- **ModelCheckpoint**: Lưu mô hình tốt nhất dựa trên validation loss.
- **LambdaCallback**: Trực quan hóa quá trình huấn luyện.

4. Cấu hình huấn luyện:

- Số epoch tối đa: 156 (xác định từ thí nghiệm trước đó).

- Batch size: 128.

3. Kết quả thí nghiệm

1. Hiệu suất với các kích thước mẫu huấn luyện:

- MSE trên tập xác thực giảm khi tăng kích thước mẫu huấn luyện, nhưng giảm dần ở mức chậm hơn khi kích thước mẫu vượt **1400**.
- Giá trị MSE nhỏ nhất: **0.000971**, đạt được với **1400 mẫu**.

2. Phân tích biểu đồ:

- Biểu đồ MSE so với kích thước mẫu cho thấy:
 - Giảm nhanh trong khoảng từ 200 đến 1400 mẫu.
 - Sau 1400 mẫu, cải thiện trở nên ít đáng kể, cho thấy mô hình đã đạt khả năng tốt nhất.

4. Phân tích kết quả

1. Hiệu suất tổng quan:

- Mô hình MLP cho hiệu suất tốt, với MSE nhỏ (<0.001) ngay cả khi kích thước mẫu huấn luyện tương đối nhỏ.
- Hiệu suất đạt tối ưu khi số mẫu huấn luyện là 1400, phản ánh khả năng học tốt của mô hình với tập dữ liệu đầy đủ thông tin.

2. Tác động của kích thước mẫu:

- Kích thước mẫu tăng giúp mô hình học tốt hơn, do dữ liệu phong phú hơn để huấn luyện.
- Tuy nhiên, việc tăng kích thước mẫu sau 1400 không cải thiện đáng kể, cho thấy mức độ dữ liệu đã đủ để mô hình đạt hiệu suất tối ưu.

5. Kết luận

- Thí nghiệm đã chỉ ra rằng kích thước mẫu huấn luyện là một yếu tố quan trọng ảnh hưởng đến hiệu suất mô hình.
- Kích thước tối ưu để đạt MSE nhỏ nhất là 1400 mẫu.
- Mô hình MLP được huấn luyện thành công, đạt hiệu suất tốt trên bài toán xấp xỉ hàm số, với MSE dưới 0.001.