

## 1. Mô tả bài toán

Bài toán hồi quy nhằm dự đoán giá nhà tại bang California dựa trên các đặc trưng đầu vào như mật độ dân cư, thu nhập trung bình, số phòng trung bình,... Sử dụng bộ dữ liệu California Housing từ `sklearn.datasets`.

## 2. Mô hình & Cấu hình huấn luyện

Áp dụng mạng nơ-ron đa tầng (MLP) với các siêu tham số:

`hidden_size`: Số neuron mỗi lớp ẩn 64 hoặc 128

`num_layers`: Số lớp ẩn (3 hoặc 5)

`learning_rate`: Tốc độ học (0.001 hoặc 0.0005)

Mỗi cấu hình chạy lặp lại 5 lần để lấy giá trị trung bình và độ lệch chuẩn RMSE.

Thời gian huấn luyện: 100 epoch, batch size = 64, dùng Adam optimizer và StepLR.

## 3. Đánh giá hiệu suất

Hiệu suất được đo qua RMSE trên dữ liệu chuẩn hóa và dữ liệu gốc (giá nhà tính theo đơn vị nghìn đô la Mỹ). Kết quả:

Cấu hình	Hidden	Layers	LR	RMSE chuẩn hóa $\pm$ std	RMSE thực tế $\pm$ std (nghìn \$)
Config 1	64	3	0.001	$0.4519 \pm 0.0092$	$0.5215 \pm 0.0106$
Config 2	64	3	0.0005	$0.4536 \pm 0.0021$	$0.5234 \pm 0.0024$
Config 3	64	5	0.001	$0.4619 \pm 0.0191$	$0.5330 \pm 0.0220$
Config 4	64	5	0.0005	$0.4599 \pm 0.0045$	$0.5307 \pm 0.0052$
Config 5	128	3	0.001	$0.4378 \pm 0.0042$	$0.5052 \pm 0.0049$
Config 6	128	3	0.0005	$0.4563 \pm 0.0195$	$0.5265 \pm 0.0225$
Config 7	128	5	0.001	$0.4468 \pm 0.0131$	$0.5156 \pm 0.0151$
Config 8	128	5	0.0005	$0.4613 \pm 0.0252$	$0.5323 \pm 0.0291$

#### 4. Nhận xét

Cấu hình tốt nhất: Config 5 (hidden=128, layers=3, lr=0.001) cho RMSE thấp nhất cả trên dữ liệu chuẩn hóa và thực tế.

Mô hình có xu hướng hiệu quả hơn với số lớp ít hơn (3 layers), khi số lớp tăng lên 5, độ lệch chuẩn thường lớn hơn và RMSE thực tế cao hơn.

Learning rate 0.001 thường cho kết quả tốt hơn 0.0005.

#### 5. Ghi chú kỹ thuật

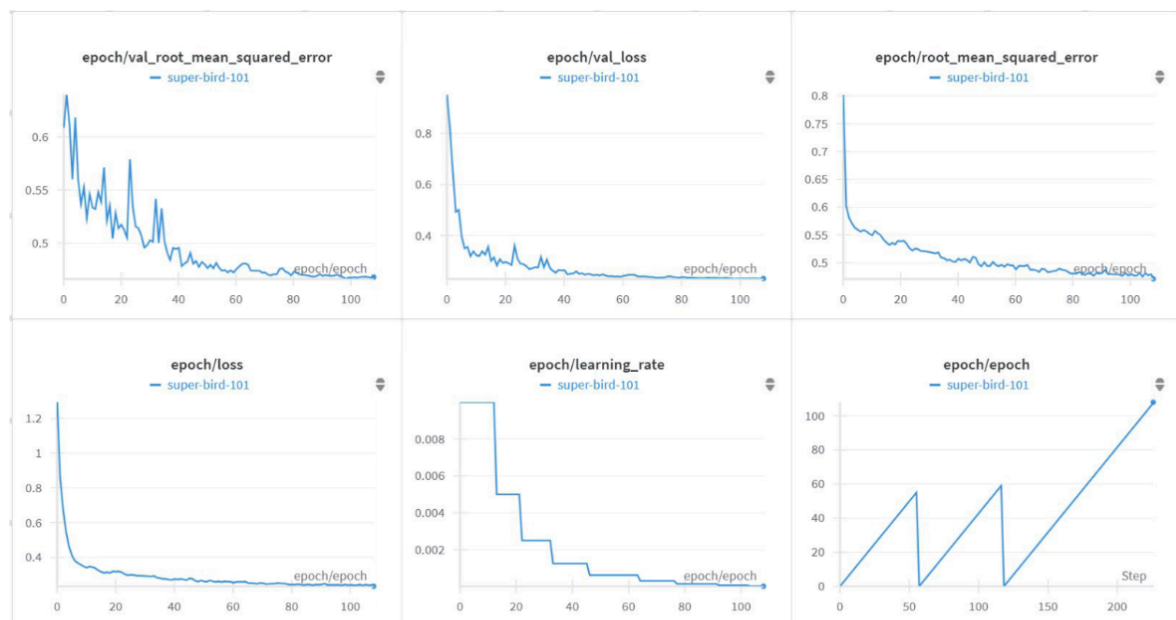
Dữ liệu không có giá trị thiếu.

Chuẩn hóa toàn bộ đầu vào và đầu ra bằng StandardScaler.

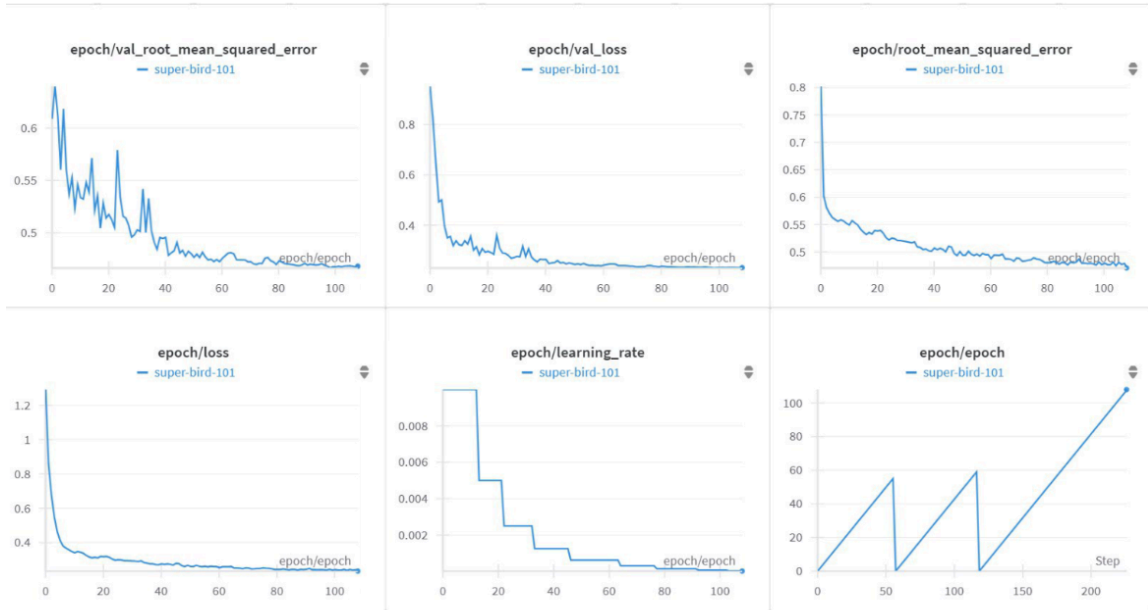
Dùng TensorBoard để theo dõi quá trình học (loss, learning rate, histogram trọng số).

#### 6. đồ thị hàm

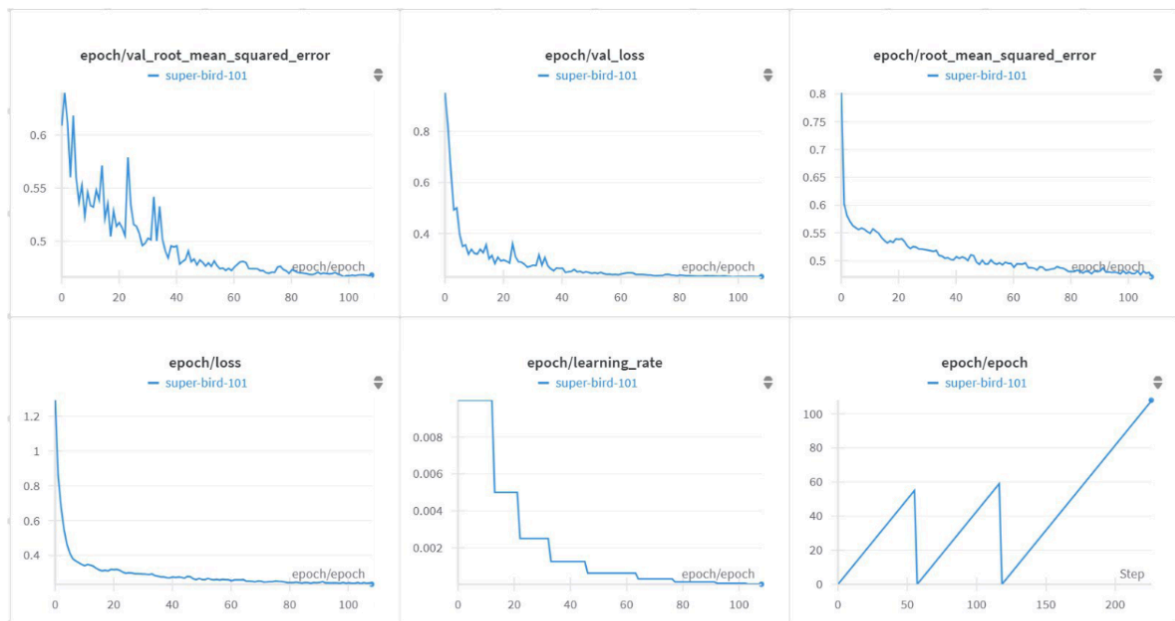
Config 1:



Config 2 :



Config 3



Config 4:

