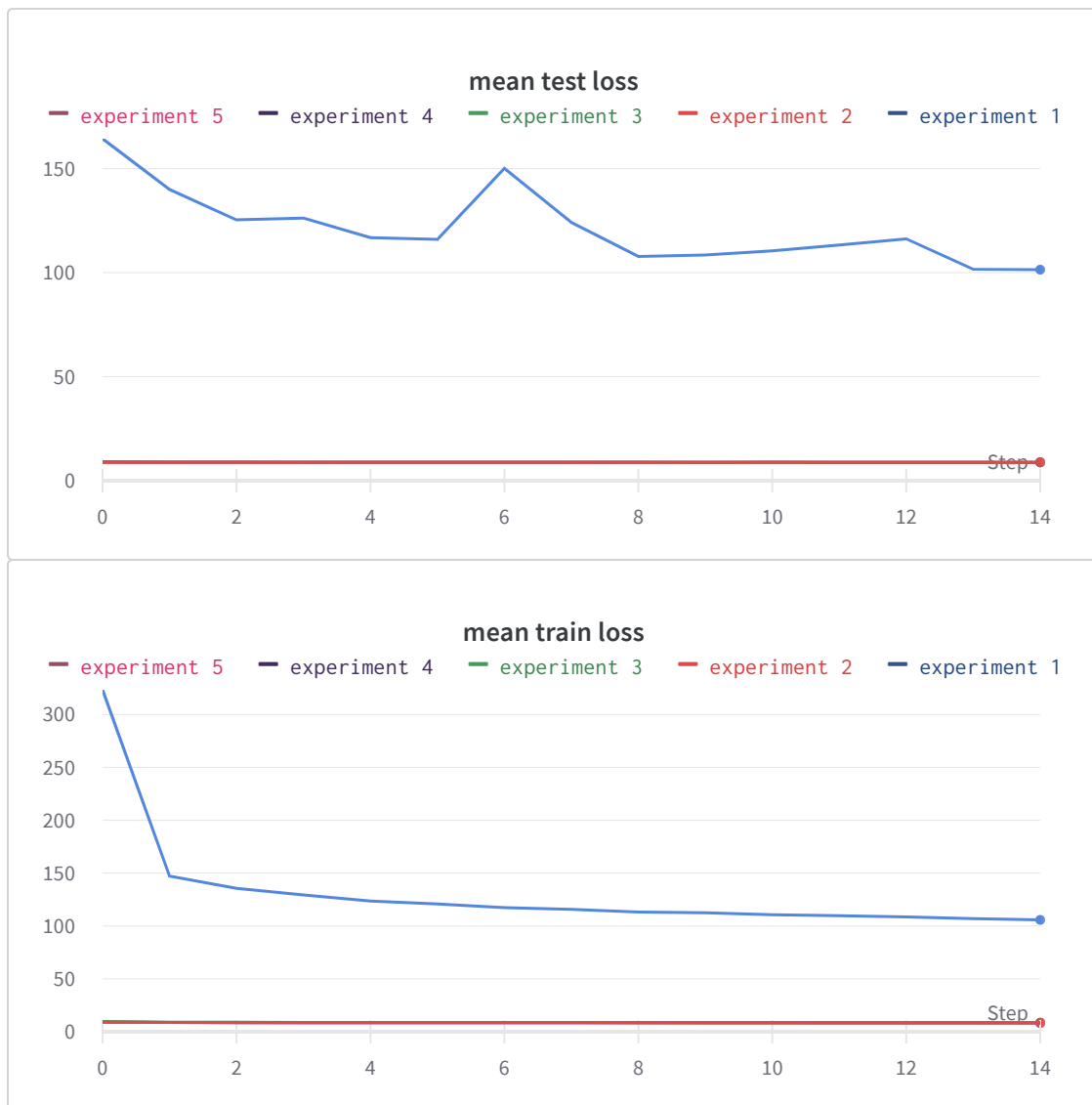


# ИАД DL ДЗ№1

Бесконечные попытки выбить необходимое качество...

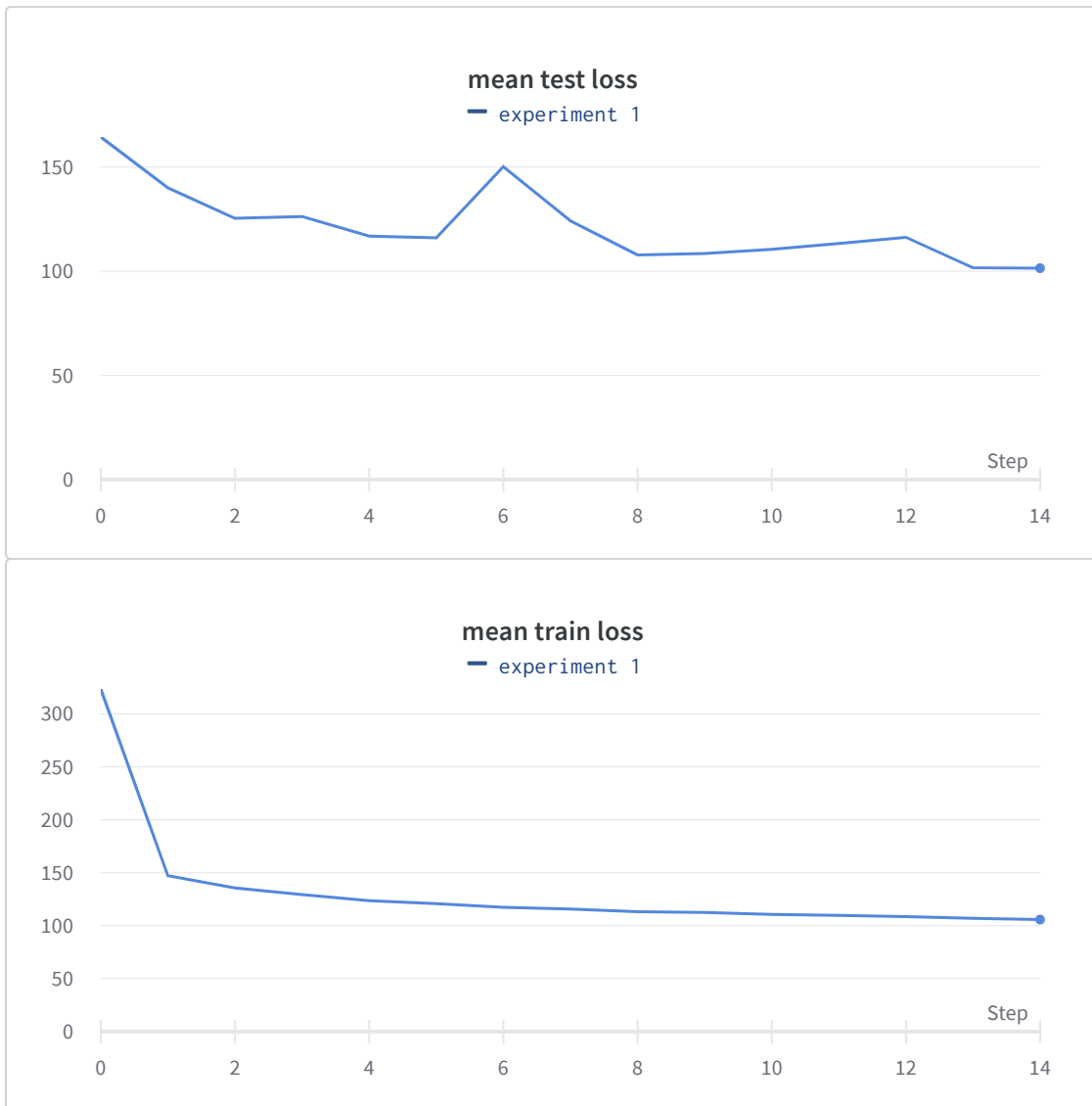
Danil Komarov

Смотря на эти два графика ниже, хочется задаться вопросом: а что это такое вообще? Создается впечатление, что вообще ничего не получилось, но это не так, давайте разбираться!



## ▼ Эксперимент 1.

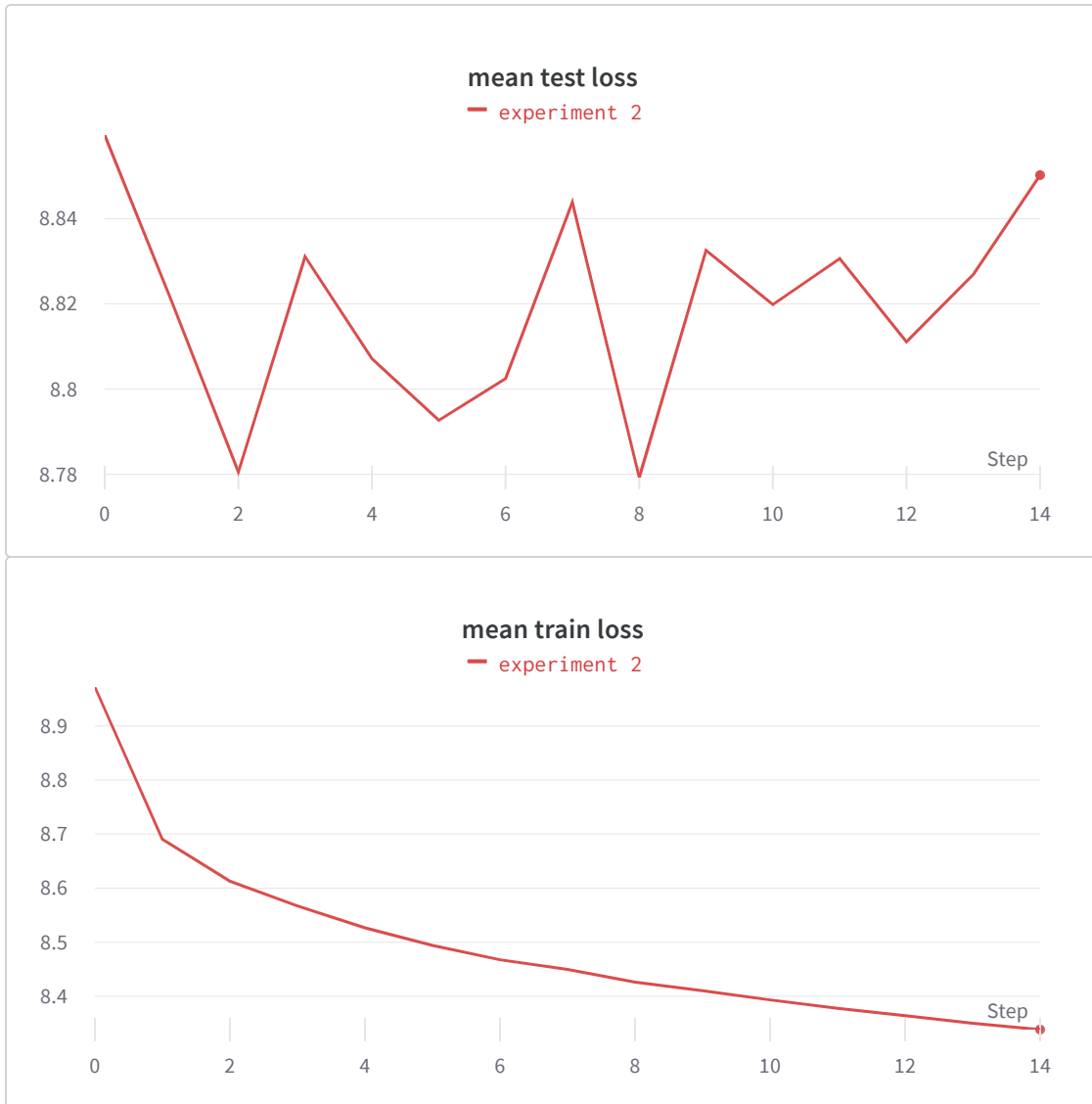
- **Ненормализованные** данные
- Модель состоит из 3 полносвязных слоев, функция активации - ReLU
- Оптимизатор Adam
- 15 эпох
- Размер батча: 256



При ненормализованных данных лосс на целый порядок выше, чем у регрессии из `sklearn`. Модель в целом училась, но это качество было очень далеко от необходимого. Дальше пошли эксперименты с нормализованными данными.

## ▼ Эксперимент 2.

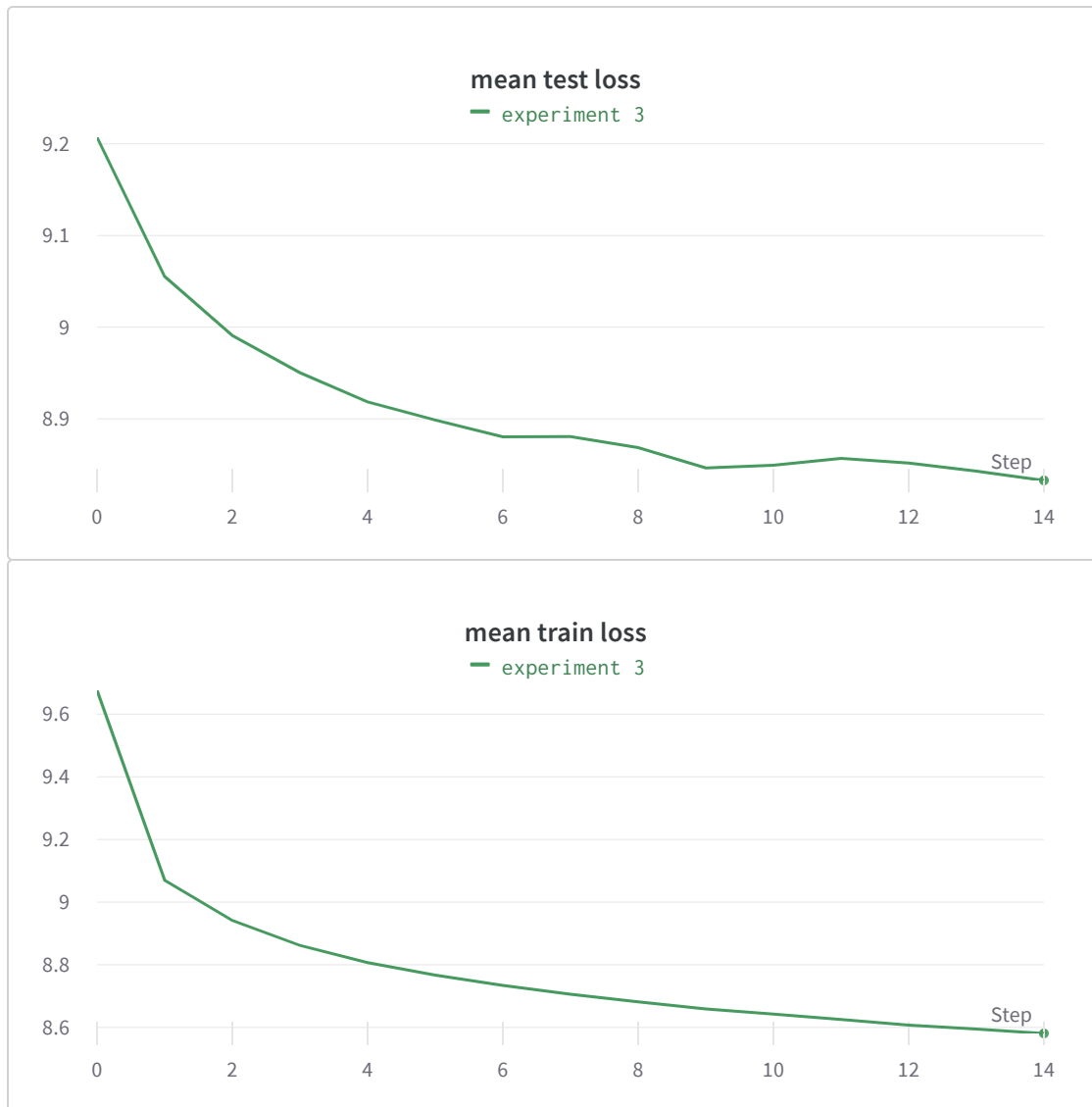
- **StandardScaler** нормализация фичей и таргета
- Модель состоит из 3 полносвязных слоев, функция активации - ReLU
- Оптимизатор Adam
- 15 эпох
- Размер батча: 256



Порядок лосса теперь то, что нужно. На отдельных эпохах видно, что качество близко к необходимому, но модель не особо учится, ведет себя от эпохе к эпохе достаточно стохастично.

## ▼ Эксперимент 3.

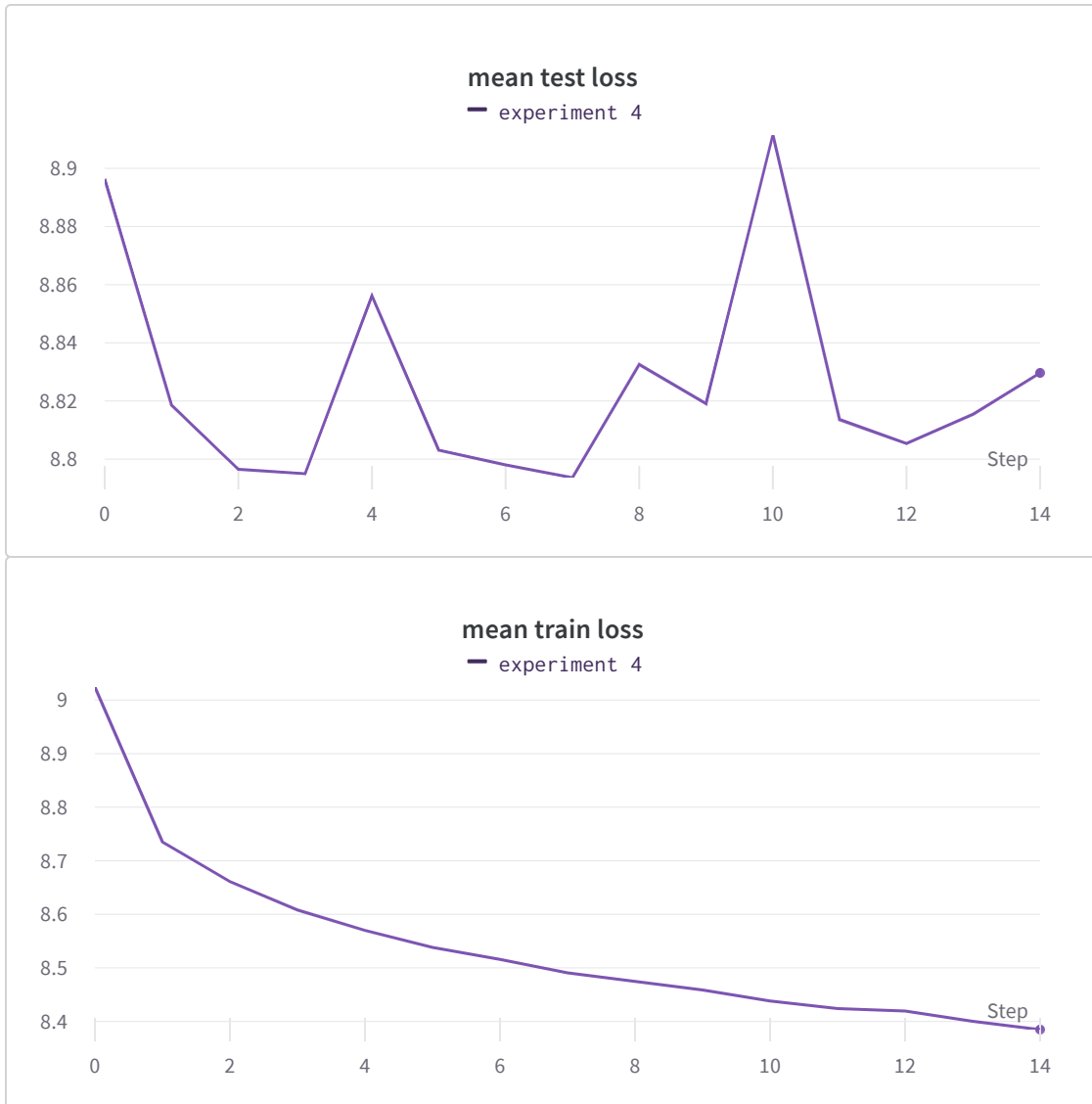
- **StandardScaler** нормализация фичей и таргета
- Модель состоит из 3 полносвязных слоев, функция активации - ReLU
- Оптимизатор SGD: lr=0.01
- 15 эпох
- Размер батча: 256



Здесь было решено попробовать другой оптимизатор - SGD. Было протестировано несколько вариантов `learning rate`, неплохим оказался 0.01, модель учится и делает это достаточно успешно. Однако она не доходит до необходимого качества, увеличение эпох не помогало, лосс начинает буксовать, поэтому последовал следующий эксперимент с этим оптимизатором.

## ▼ Эксперимент 4.

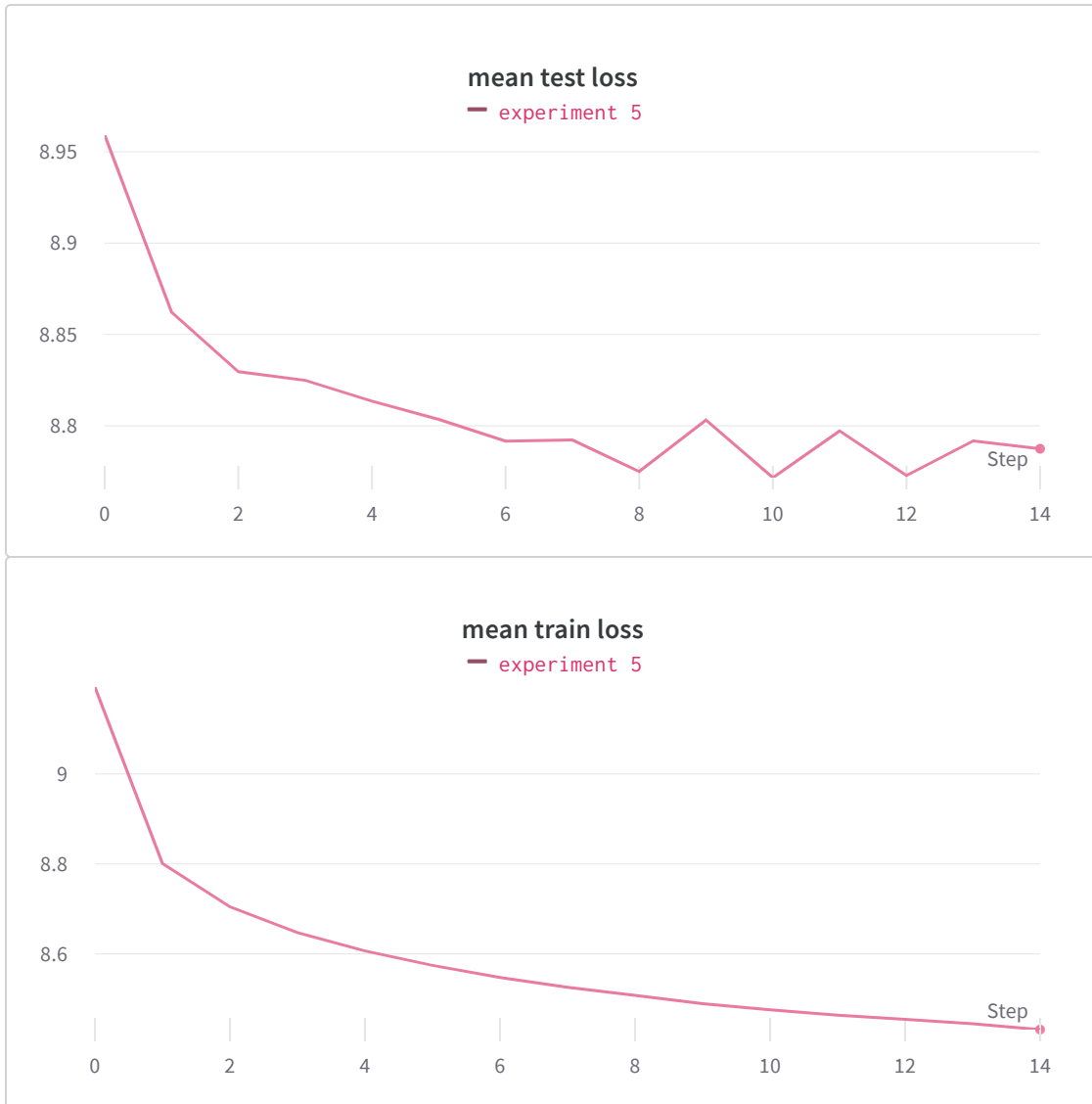
- **StandardScaler** нормализация фичей и таргета
- Модель состоит из 3 полносвязных слоев, функция активации - ReLU
- Оптимизатор SGD:  $lr=0.01$ ,  $momentum=0.9$
- 15 эпох
- Размер батча: 256



В этом эксперименте был добавлен `momentum`, было попробовано несколько вариантов, но какого-то успеха добиться не получилось, лосс на тесте начал вести себя более стохастично по сравнению с предыдущим экспериментом.

## ▼ Эксперимент 5. FINAL

- **StandardScaler** нормализация фичей и таргета
- Модель состоит из 3 полносвязных слоев, функция активации - ReLU
- Оптимизатор Adam: lr=0.0003, weight\_decay=0.0005
- 15 эпох
- Размер батча: 256



Снова вернулся к Adam. Немного поэкспериментировал с learning rate и добавил weight\_decay. Модель начала учиться очень неплохо и лосс начал радовать. Именно этот train loop был финальным. Сохраняю здесь модель с лучшим лоссом.

## ▼ Итоги

1. Количество полносвязных слоев не особо сильно меняет ситуацию, 2 или 3 разницы особой нет, если ставить больше - профита я не заметил.
2. Нормализация данных дает сильный буст качеству.
3. Dropout в моем случае никак не помог.
4. Перебор гиперпараметров у оптимизаторов дает неплохой результат.

Created with ❤ on Weights & Biases.

5. Эксперименты 3 и 5 имеют полное право на жизнь и адекватность,

[https://wandb.ai/igor\\_novikov/keras-1-conv2d-1-zsx](https://wandb.ai/igor_novikov/keras-1-conv2d-1-zsx) выбран был 5 ввиду лучшего качества на тестовой выборке.

6. RMSE на тестовой выборке: **8.775**.