

Лекция 6

bag of tricks

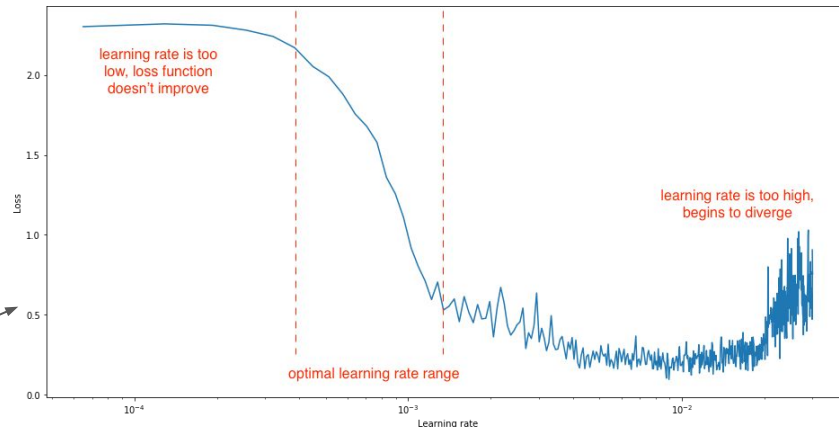
BatchNorm2D

$$\gamma \hat{x} + \beta.$$

- В случае если batch norm в res блоке
 - зануляем гамма при инициализации
 - Тем самым resnet блок в начальных итерациях будет возвращать значение подаваемое на вход, тем самым будучи похожей на маленькую сетку
- Если batchnorm идет после conv, выставляем параметр bias=False у conv
- Если в сетке применяется регуляризация на веса, она не должна применяться на параметры батч норма

Learning rate

- Делаем learning rate warmup, т.е. в начале очень маленький lr, далее увеличиваем



- OneCycleLR
 - LR range test - определяем максимальный lr который не увеличивает лосс

- [torch.optim](#)

Label Smoothing

- В задачах классификации обычно установлены hard targets, т.е. Строго указан один из классов
- Это может привести к переобучению
- Один из способ избежать этого, задавать таргеты следующим образом:

$$q_i = \begin{cases} 1 - \varepsilon & \text{if } i = y, \\ \varepsilon / (K - 1) & \text{otherwise,} \end{cases}$$

Дополнительно

- В случае если решаем задачу регрессии, то инициализируем bias последнего линейного слоя средним значением на выборке, тем самым увеличиваем скорость сходимости
- Не спешите останавливать модель, если не видите улучшения метрики пару эпох