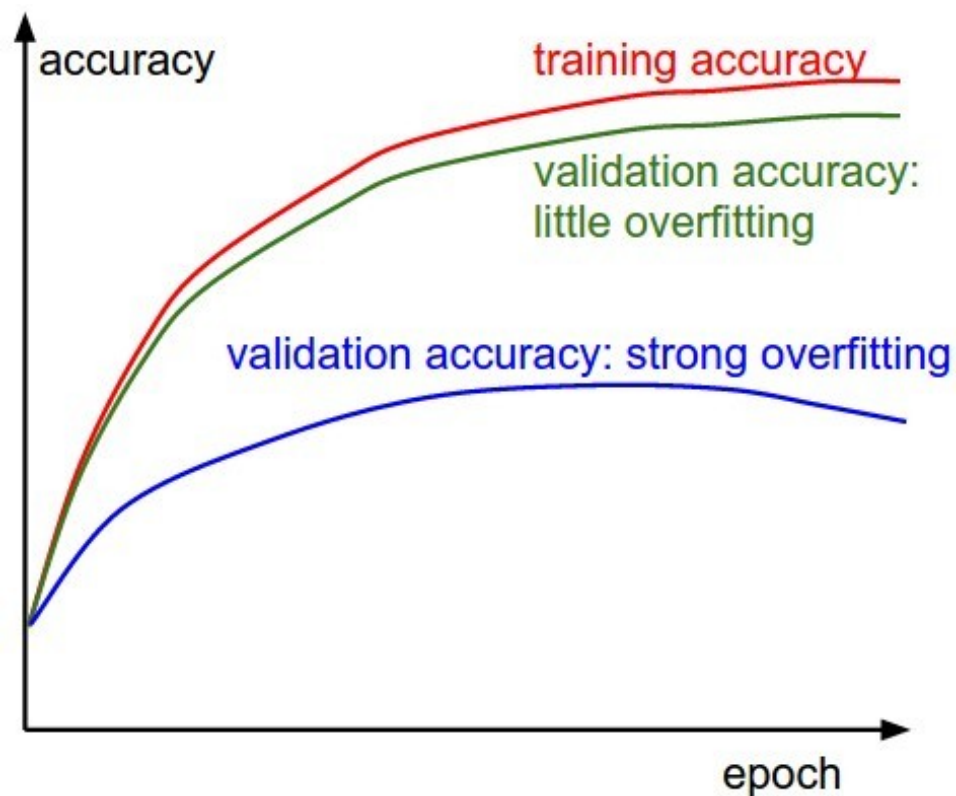
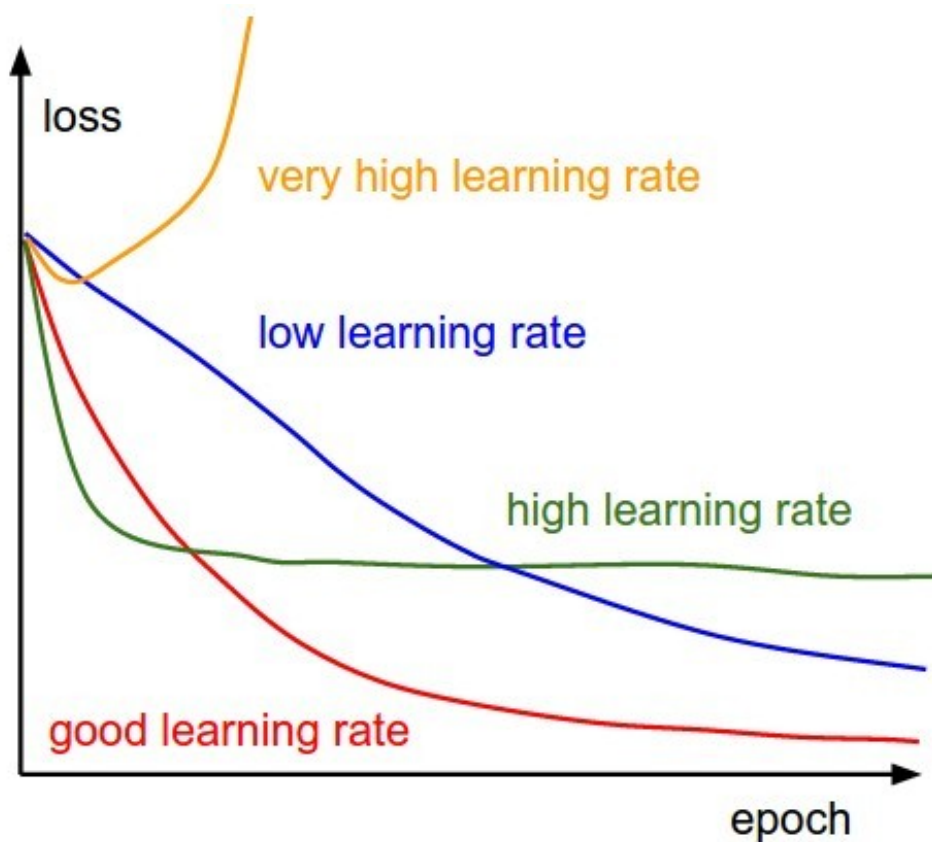


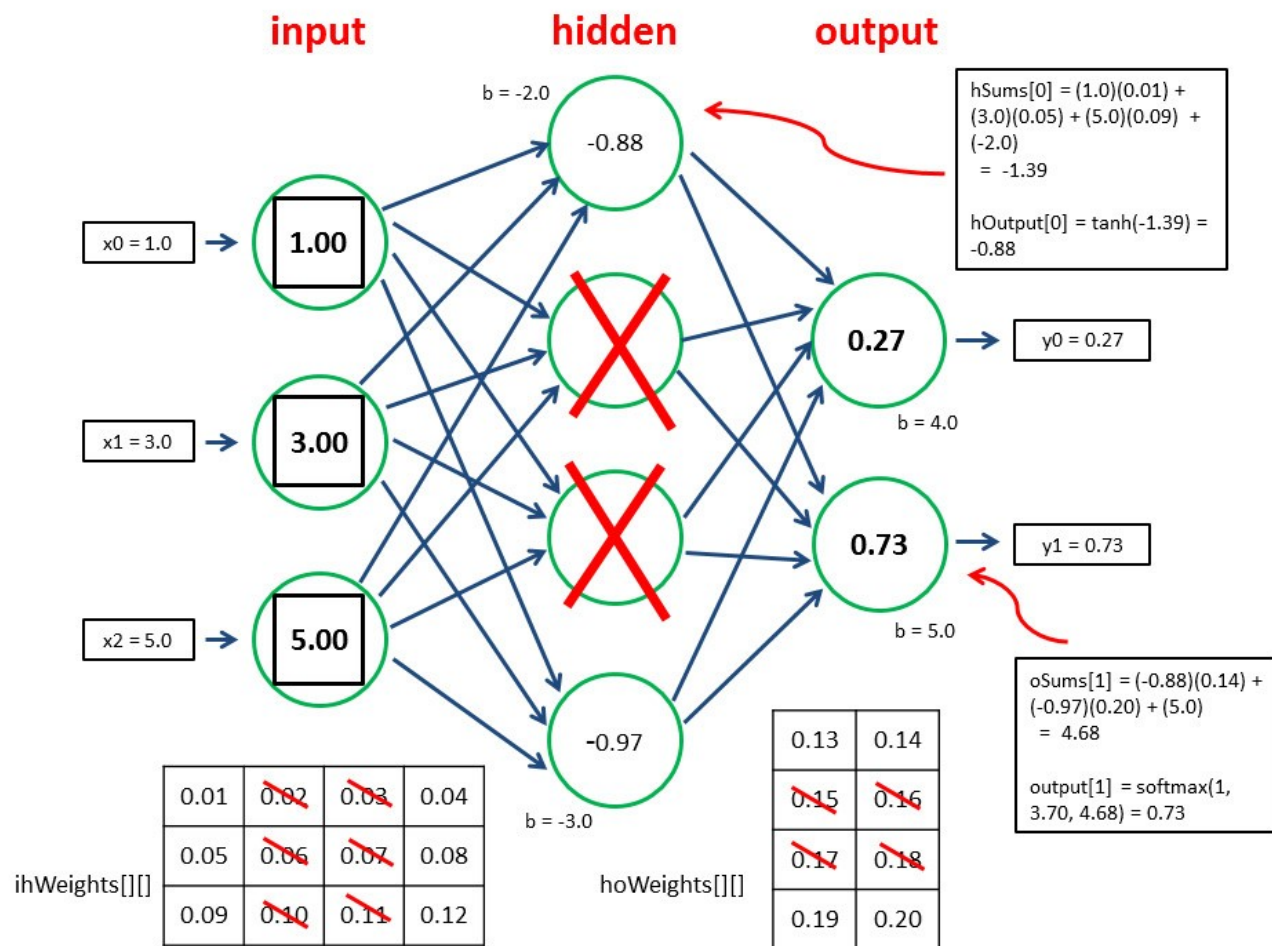
Глубокое обучение для обработки изображений

Лекция 4

Типичная зависимость лосса и метрик от номера эпохи



Слои CNN: Dropout



Dropout применяется для регуляризации, а также уменьшения влияния одних нейронов на другие

На этапе предикта используются все нейроны

Слои CNN: Dropout

Dropout

```
class torch.nn.Dropout(p=0.5, inplace=False)
```

[\[source\]](#)

During training, randomly zeroes some of the elements of the input tensor with probability p .

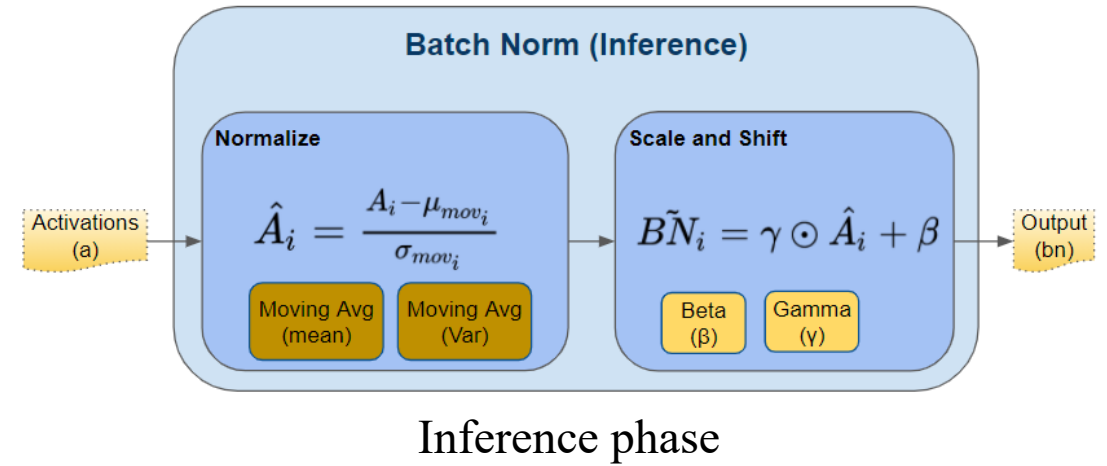
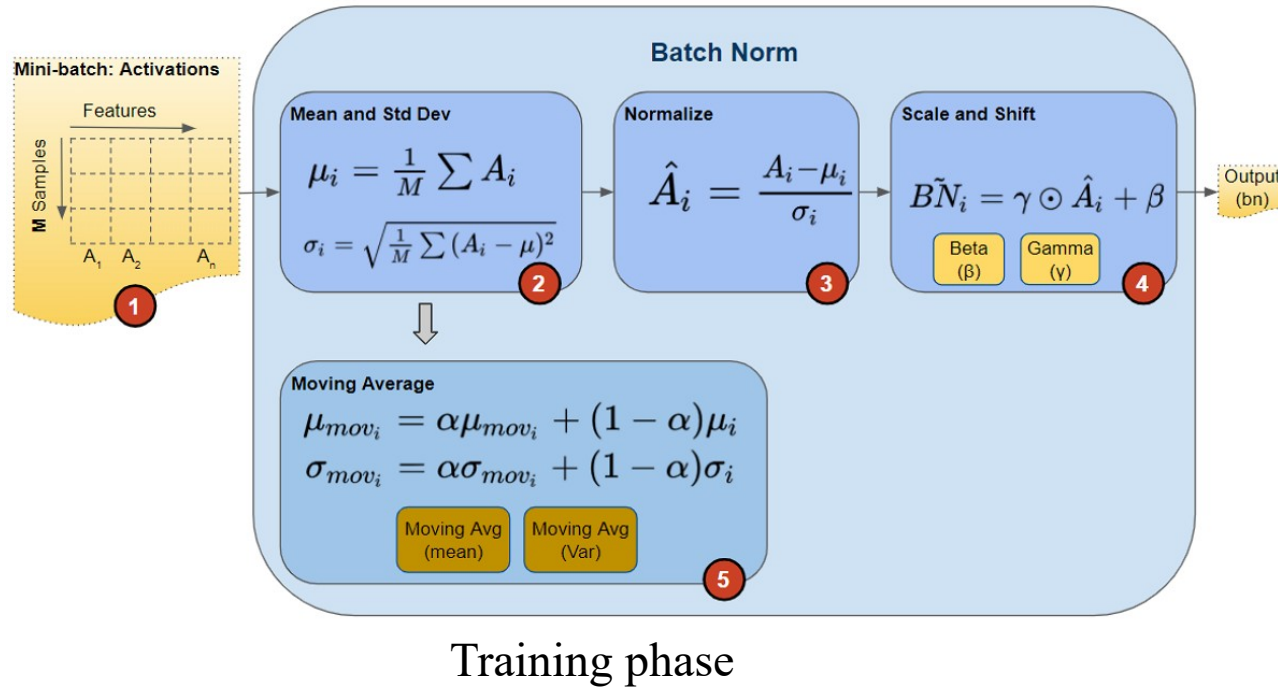
The zeroed elements are chosen independently for each forward call and are sampled from a Bernoulli distribution.

Each channel will be zeroed out independently on every forward call.

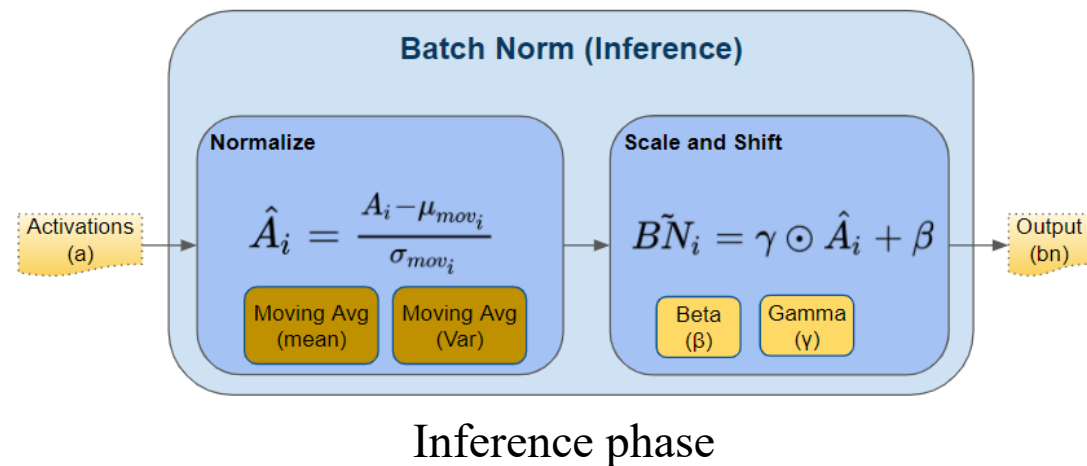
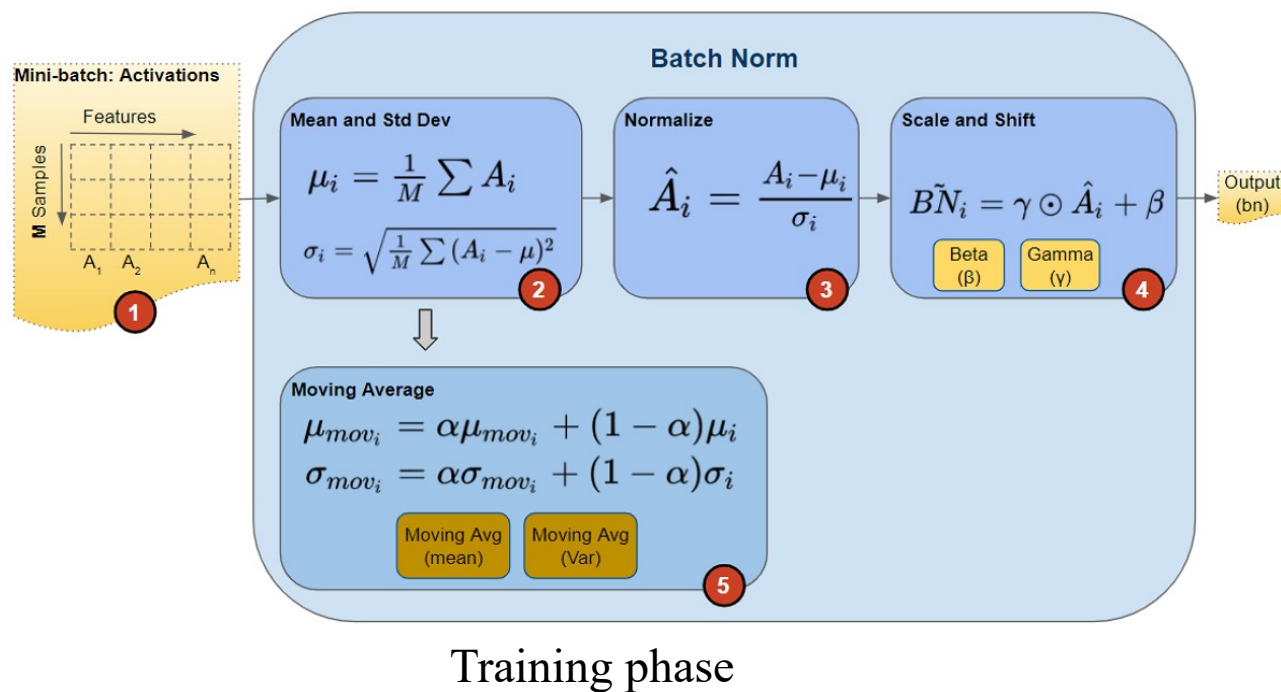
This has proven to be an effective technique for regularization and preventing the co-adaptation of neurons as described in the paper [Improving neural networks by preventing co-adaptation of feature detectors](#).

Furthermore, the outputs are scaled by a factor of $\frac{1}{1-p}$ during training. This means that during evaluation the module simply computes an identity function.

Слои CNN: Batch Normalization



Слои CNN: Batch Normalization



- Более стабильное обучение: значения приводятся к одному диапазону → меньше вероятность затухания/взрыва градиента
- Борется с internal covariate shift (распределение входных данных слоя постоянно меняется в процессе обучения), который усложняет и замедляет сходимость
- Позволяет использовать большие значения LR

Слои CNN: Batch Normalization

BatchNorm2d

```
class torch.nn.BatchNorm2d(num_features, eps=1e-05, momentum=0.1, affine=True,  
track_running_stats=True, device=None, dtype=None)
```

[\[source\]](#)

Applies Batch Normalization over a 4D input.

4D is a mini-batch of 2D inputs with additional channel dimension. Method described in the paper [Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift](#).

$$y = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \epsilon}} * \gamma + \beta$$

The mean and standard-deviation are calculated per-dimension over the mini-batches and γ and β are learnable parameter vectors of size C (where C is the input size). By default, the elements of γ are set to 1 and the elements of β are set to 0. At train time in the forward pass, the standard-deviation is calculated via the biased estimator, equivalent to `torch.var(input, unbiased=False)`. However, the value stored in the moving average of the standard-deviation is calculated via the unbiased estimator, equivalent to `torch.var(input, unbiased=True)`.

Also by default, during training this layer keeps running estimates of its computed mean and variance, which are then used for normalization during evaluation. The running estimates are kept with a default `momentum` of 0.1.

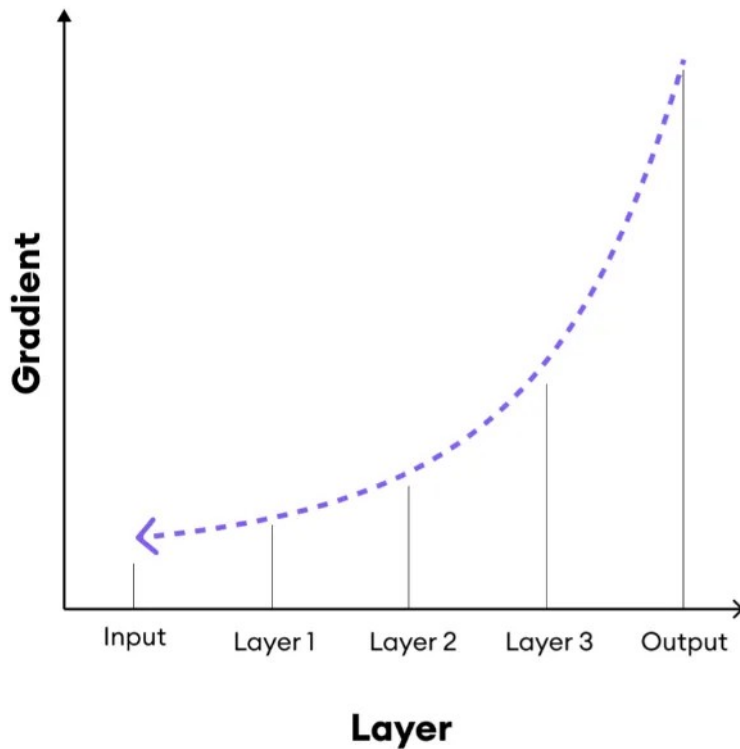
If `track_running_stats` is set to `False`, this layer then does not keep running estimates, and batch statistics are instead used during evaluation time as well.

Shape:

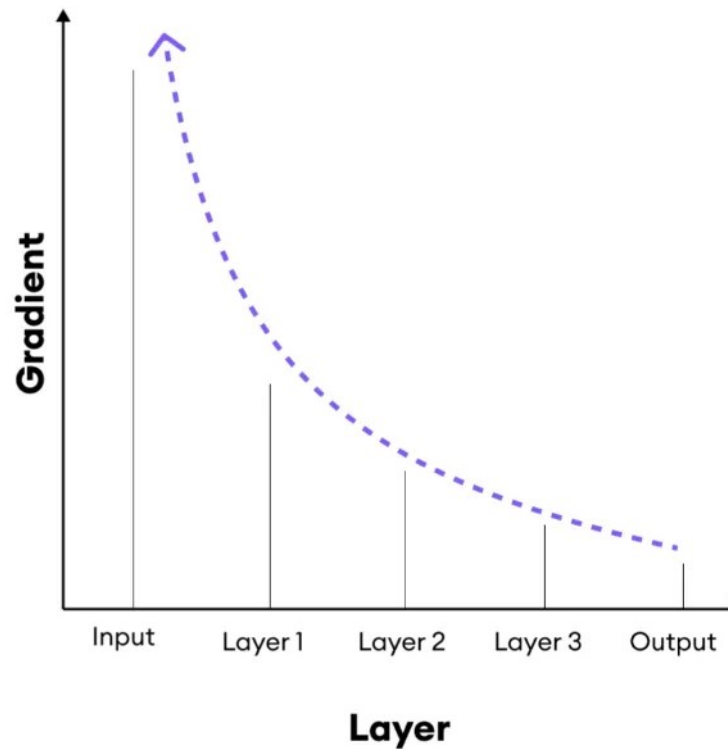
- Input: (N, C, H, W)
- Output: (N, C, H, W) (same shape as input)

Проблемы: взрывающийся (exploding) и затухающий (vanishing) градиент

Vanishing Gradient



Exploding Gradient



Некоторые методы борьбы:

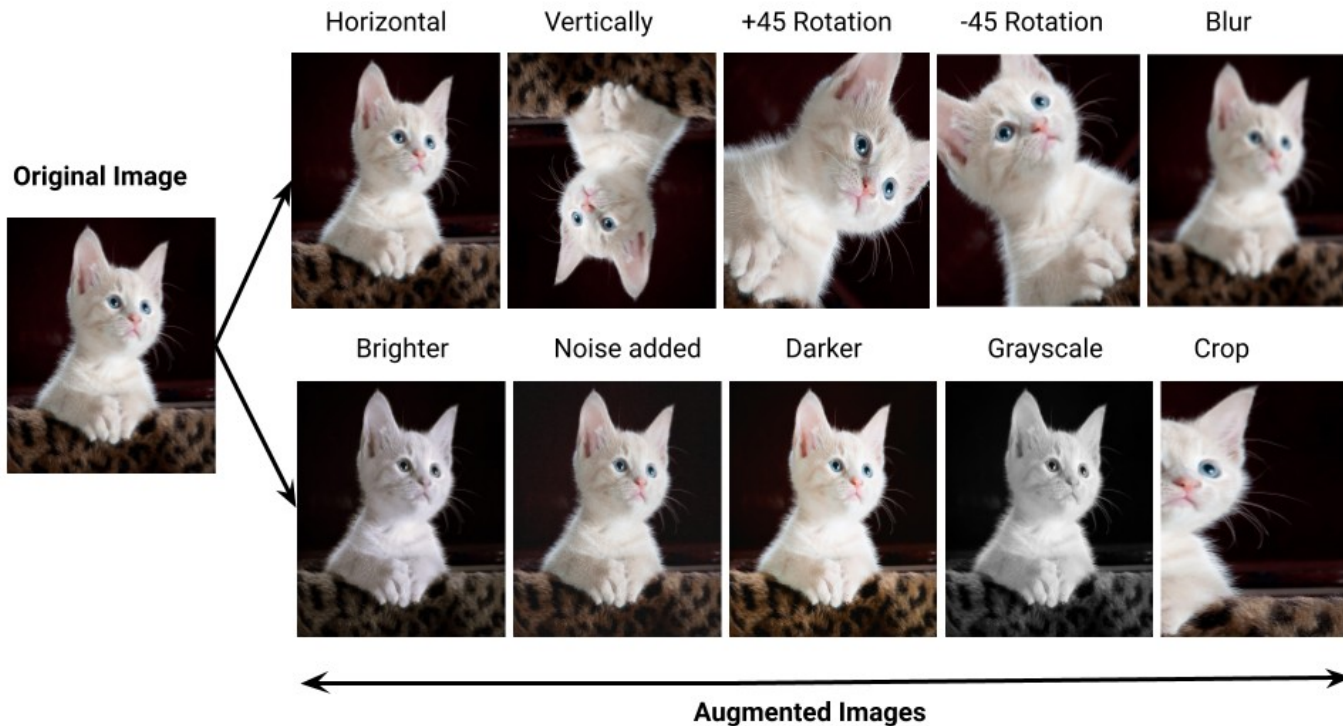
С исчезающим градиентом:

- Skip connections
- Инициализация весов (He/Xavier)
- Разновидности ReLU (чтобы в отрицательной области градиент не был равен 0)

С взрывающимся градиентом:

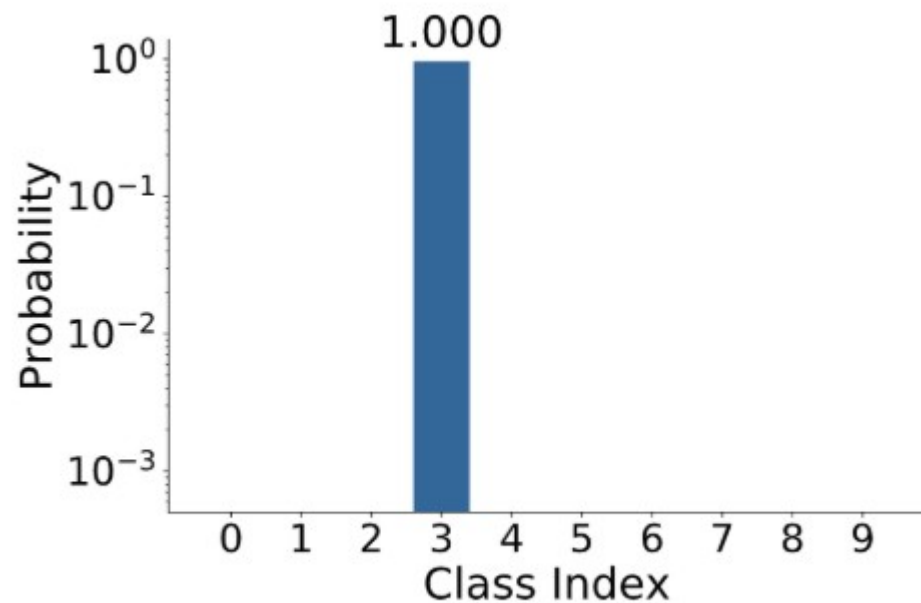
- Gradient clipping
- Инициализация весов
- BatchNorm, LayerNorm etc.

Аугментации

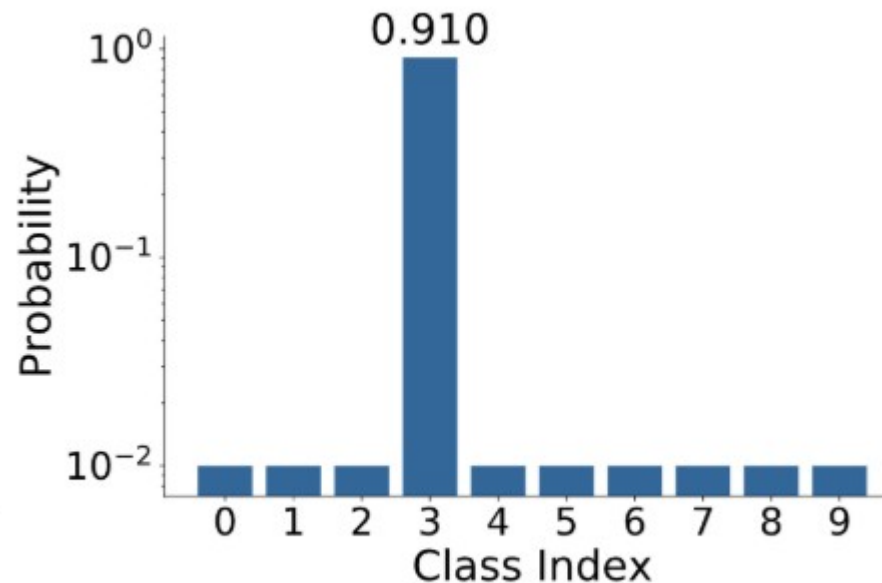


Data augmentation позволяет искусственно расширить имеющийся датасет путем применения некоторых преобразований к изображениям на каждой эпохе. Позволяет бороться с переобучением

Label smoothing



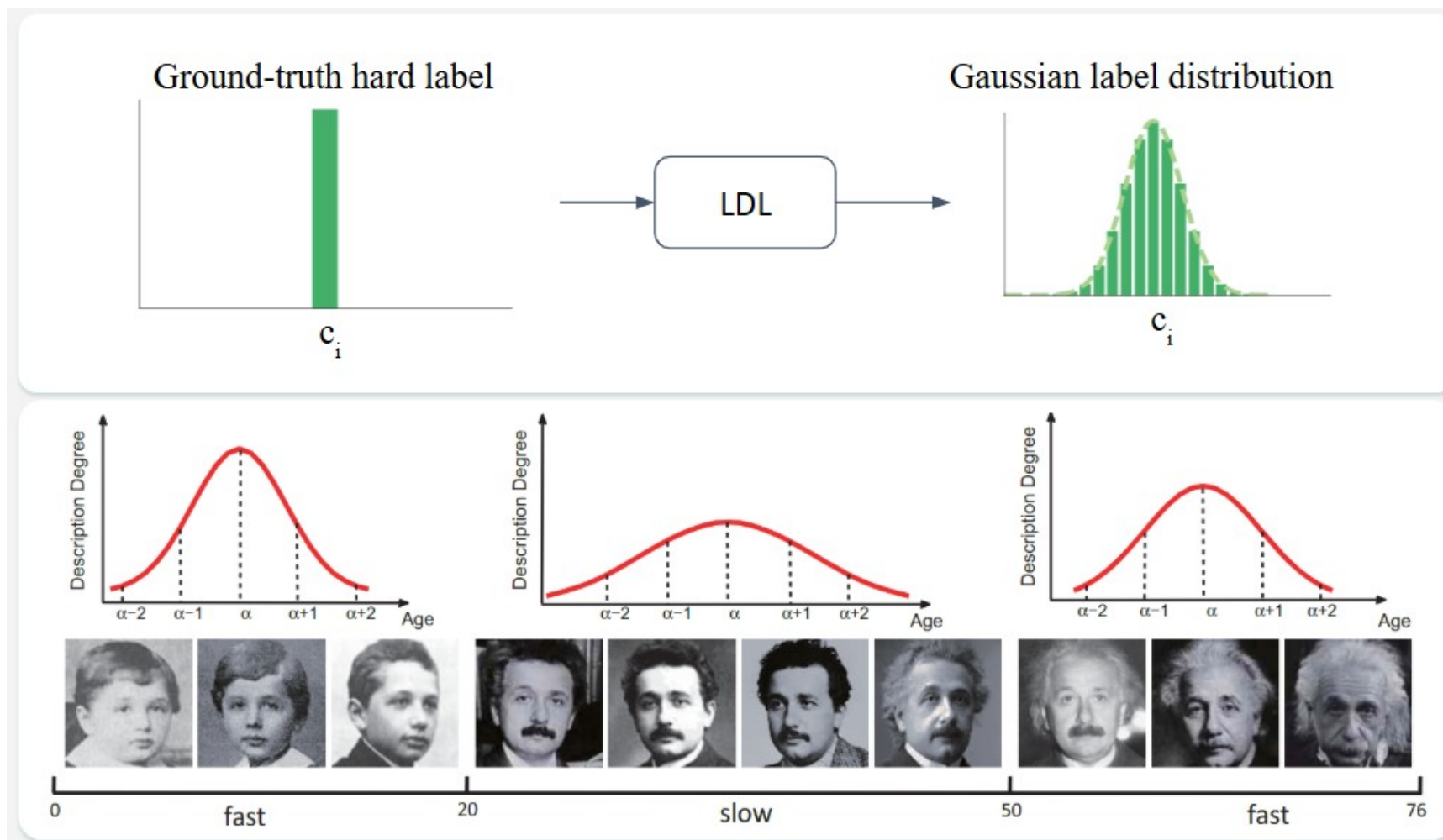
(a) Hard Label



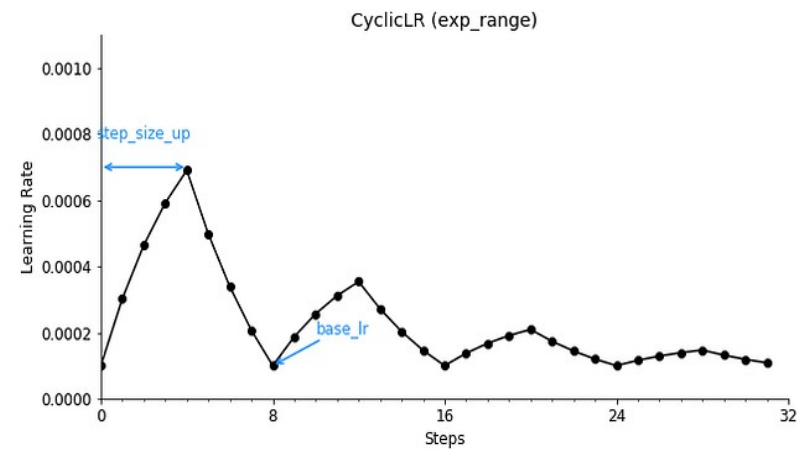
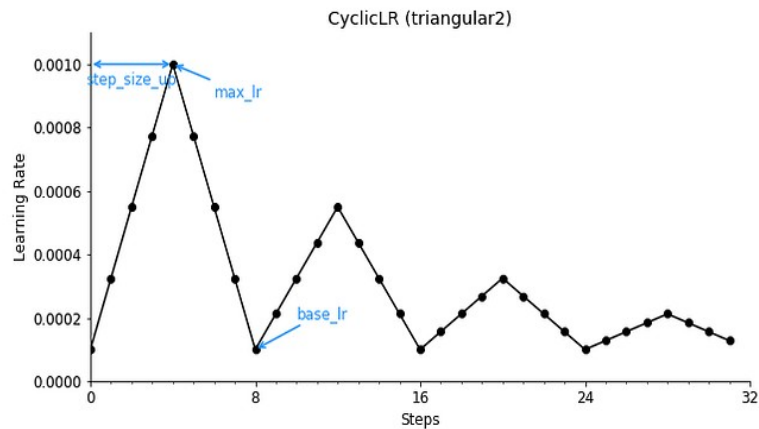
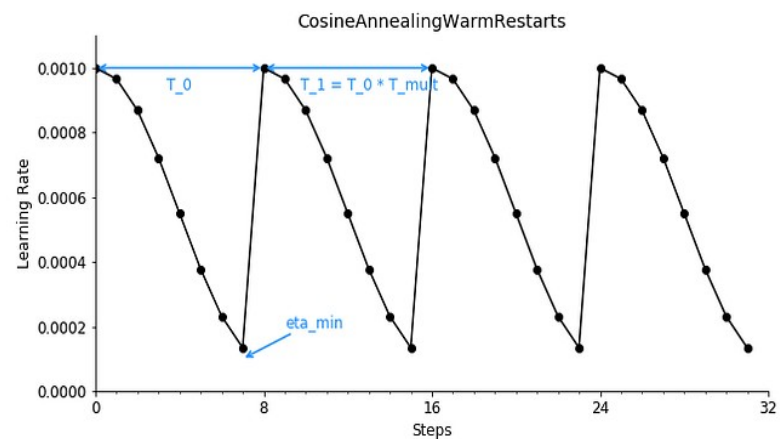
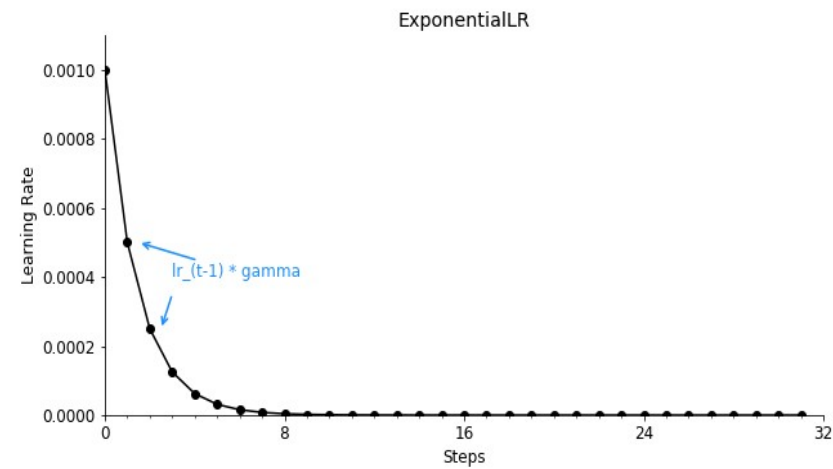
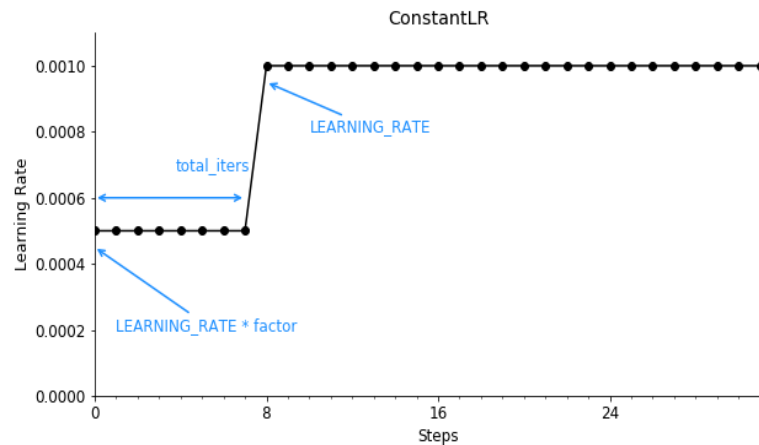
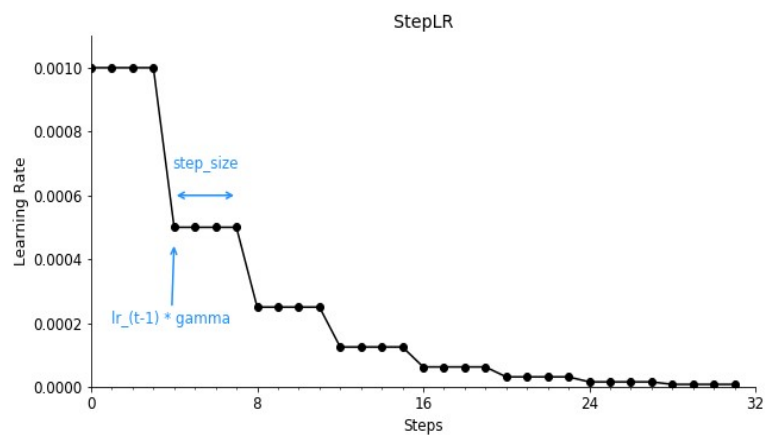
(b) LS

Сглаживание меток — распространенная техника регуляризации, которая предотвращает overconfidence и улучшает обобщение (generalization) модели

Label Distribution Learning



Learning Rate scheduler



Transfer learning (перенос обучения) и fine tuning

