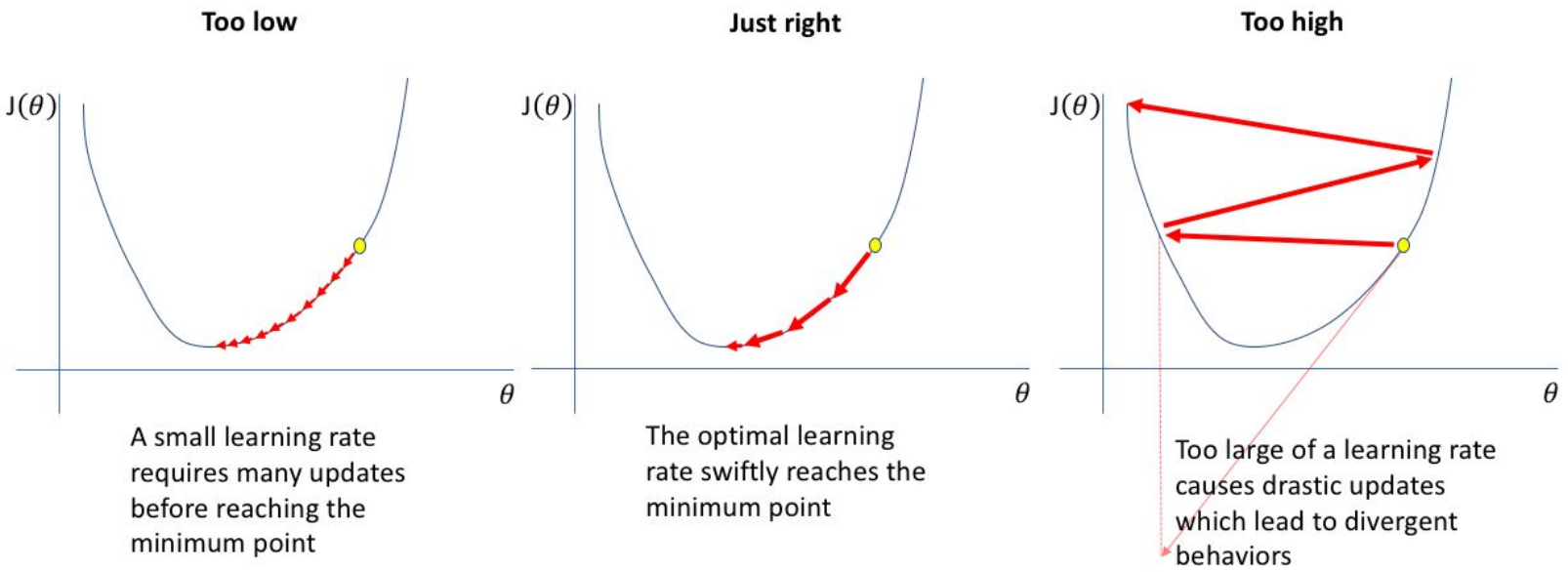
Эффективное обучение нейросетей

Основы Deep Learning

Learning rate

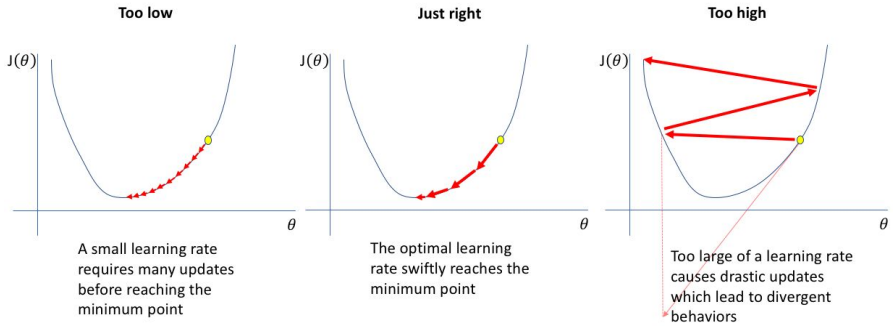
$$w^{j+1} = w^j - \alpha \frac{\partial Loss}{\partial w}(w^j)$$

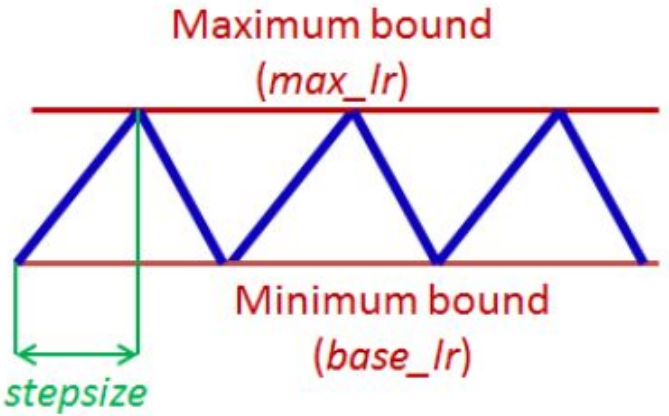
Learning rate schedule



https://arxiv.org/abs/1506.01186

Cyclic learning rate





https://arxiv.org/abs/1506.01186

Initialization

- Нулями
- Не нулями, но одинаковыми числами
- Случайными числами
- He init / Xavier init
- Новая статья: https://arxiv.org/pdf/1704.08863.pdf

Weight decay

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2.$$

L1 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda\sum_{j=0}^{M}|W_j|$$

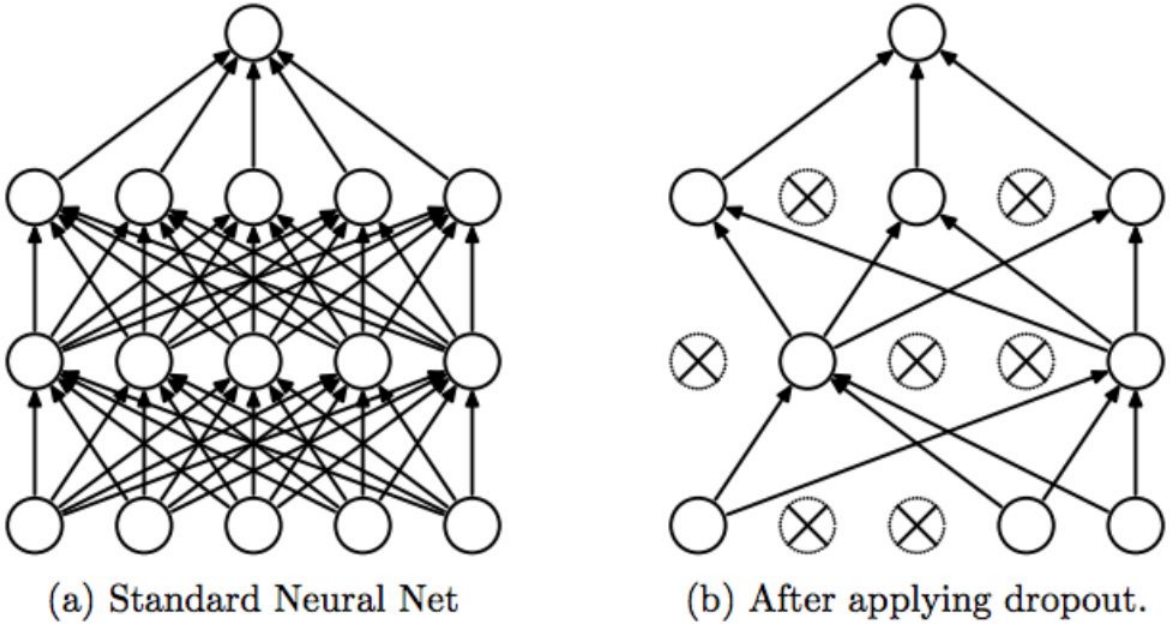
L2 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda\sum_{j=0}^{M}W_j^2$$

Loss function

Regularization Term

Dropout



(a) Standard Neural Net

(b) After applying dropout.