

极客大学机器学习训练营 深度学习

王然

众微科技 AI Lab 负责人

二〇二一年三月六日

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
- 5 参考文献

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
- 5 参考文献

- ▶ 神经网络建模过程；
- ▶ 神经网络常见元素；
- ▶ 神经网络常见设计模式。

- ▶ 随机函数与再现性；
- ▶ 验证集和模型选择。

- ▶ 在大部分时候，神经网络和人脑的神经构成有很大区别；
- ▶ 相关文章一般会习惯性的将一些设计模式赋予生物学意义；
- ▶ 还有一些文章将一些生物学的发现作为 motivation；
- ▶ 但是从生物学出发并不意味着神经网络就一定比其他模型更优越，这一点在表格化数据当中非常明显。

- ▶ 大部分时候，神经网络的学习只能通过阅读 paper（不建议阅读二手材料）和复现的方式；
- ▶ 在本章教学中，我们尽可能模仿现实学习和工作中读 paper 的状态讲授内容。

- ▶ (大部分时候) 阅读 paper 面临的问题是读不过来而不是没得读的问题;
- ▶ 很多 paper 的结果很难复现;
- ▶ 寻找有效 paper 的主要途径: Survey 和高规格比赛;
- ▶ 将一篇知名 paper 研究透彻比随意浏览若干篇有更大帮助。

- ▶ 摘要；
- ▶ 实验结果：用于判断 paper 的实际效果；
- ▶ 方法论 + 代码：注意观察不太常见的设计、注意很多实现可能有错误。

- ▶ 复现的原因：
 - ▶ 找到 paper 当中实际有效和无效的部分；
 - ▶ 纠正错误的实现；
 - ▶ 对 paper 进行改进。
- ▶ 大部分时候，不要自己从头来写。

- ▶ 神经网络的训练过程比传统模型要更复杂；
- ▶ 本章只会对整体过程做简单介绍；
- ▶ 更复杂的训练过程将会在第十一章讲解。

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
- 5 参考文献

见第四章复习

- ▶ 建模者很容易调整模型的数学形式；
- ▶ 由于目标从构建始终有效的模型变为只需要对当前数据集有效的模型，这使得神经网络的架构创新变得十分重要。

- ▶ 由于后向传导的存在，使得如果某一层出现梯度爆炸和梯度消失的现象，那么很可能之前的层也有遇到同样的问题；
- ▶ 这是神经网络设计模式当中需要解决的最重要的问题之一。

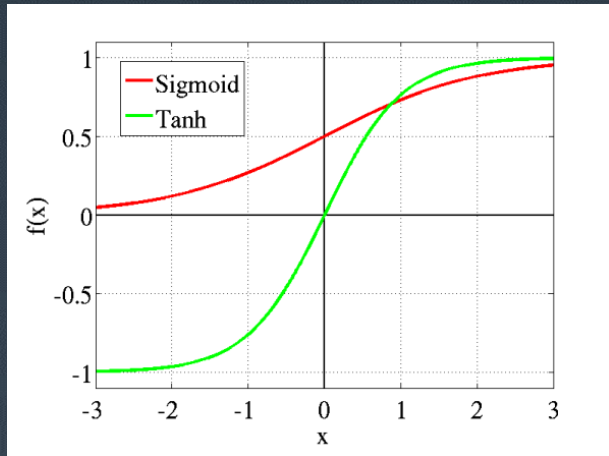
- ▶ 一般采用随机初始化；
- ▶ 分布：均匀分布和正态分布；
- ▶ He Initialization (He et al. 2015)；
- ▶ 一般来说，如果怀疑初始化有问题，只能调试方差。

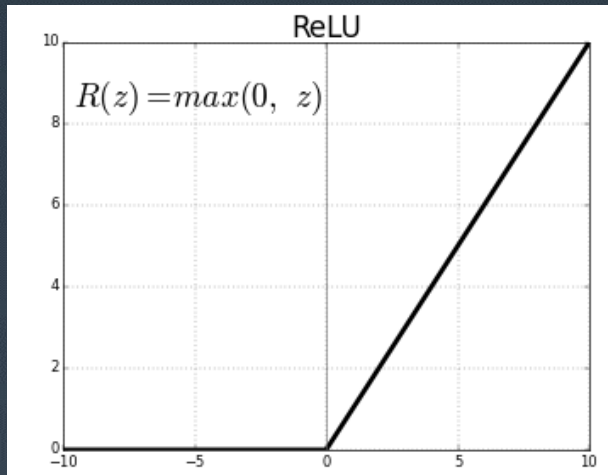
- ▶ SGD 和 SGD+Momentum;
- ▶ Adagrad 和 Adam;
- ▶ 如果不需要精细调参, 一般来说可采用 Adam;
- ▶ 对于卷积神经网络 (CV 应用), SGD+Momentum 常常效果更好。

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
- 5 参考文献

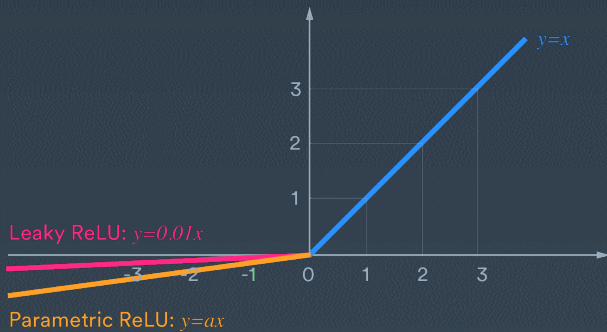
- ▶ 由于线性函数的组合仍然为线性函数，所以一般来说需要进行非线性变换；
- ▶ 问题：什么样的非线性变换是合适的？

sigmoid 和 tanh 是否适合作为激活函数？



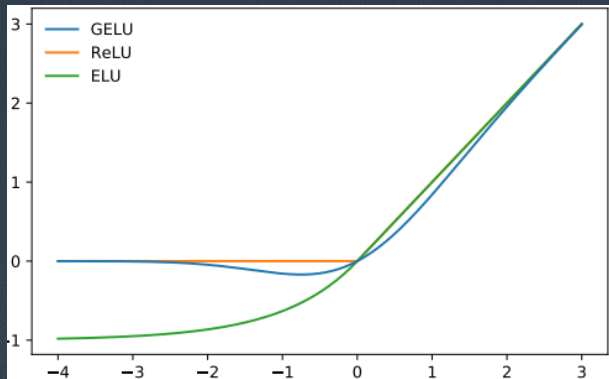


Leaky ReLU 和 PReLU

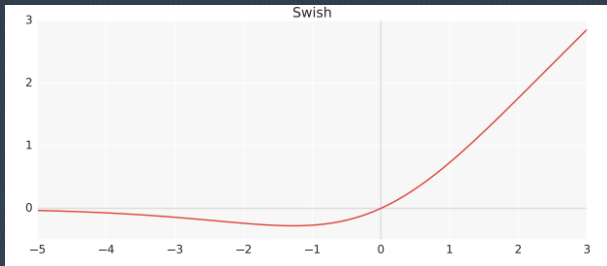


- ▶ ELU: 在 $x \geq 0$ 时和 ReLU 一样, 当 $x < 0$ 时为 $\alpha(e^{\alpha x} - 1)$;
- ▶ GeLU: $f(x) = x\Phi(x)$, 其中 $\Phi(x)$ 为正态分布的 CDF。

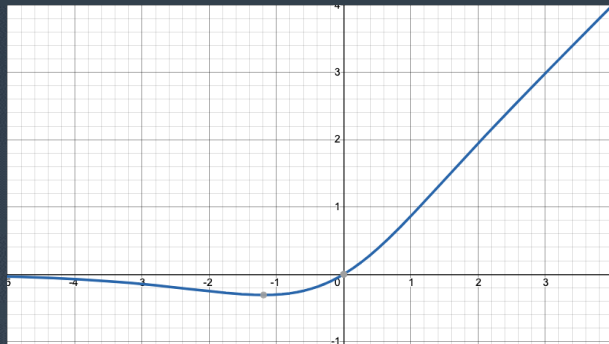
ELU 和 GeLU



数学形式: $x\sigma(x)$



数学形式: $x \tanh(\log(1 + e^x))$



- ▶ 在很多时候，使用什么激活函数一般是由预训练模型决定的；
- ▶ 并没有任何研究表明一种激活函数比另外一种激活函数系统性更优；
- ▶ 如果从头训练，建议尝试使用不同的激活函数。

- ▶ 为简化问题，考虑 ReLU 激活函数；
- ▶ 我们希望输入应该在 0 周围；
- ▶ 我们希望输入不应该太大或太小；
- ▶ 直接方法：标准化
- ▶ 问题：直接进行标准化会出现什么问题？

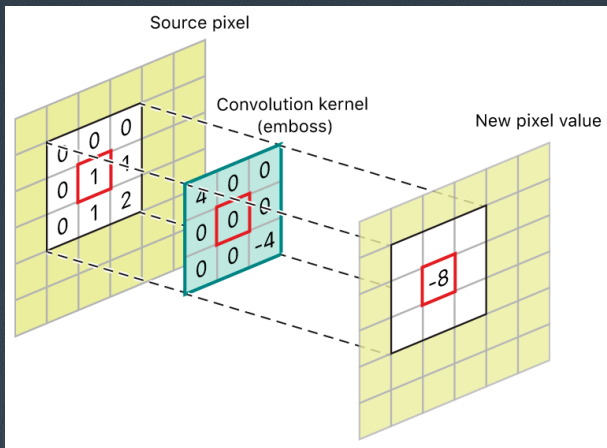
- ▶ 设定一个真实的位置和方差参数，并进行估计；
- ▶ 好处：在需要还原的时候，可以将输出进行还原；
- ▶ 数学形式： $\hat{x}_i = \gamma \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} + \beta$

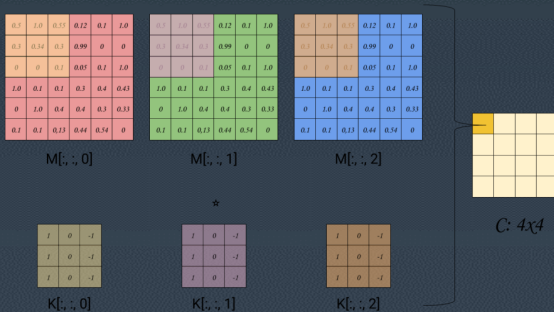
- ▶ 既可以使用在激活函数前，也可以使用在激活函数后；
- ▶ Layer Normalization, Instance Normalization 和 Group Normalization；
- ▶ Ghost Batch Normalization；
- ▶ 训练和推断不一致的解决方法；
- ▶ 详见Summers and Dinneen (2019)。

- ▶ 思想：每次随机丢弃一些神经元；
- ▶ 可以在 Normalization 之前或者之后；
- ▶ 效果有限，过大或过小都会导致训练无效。

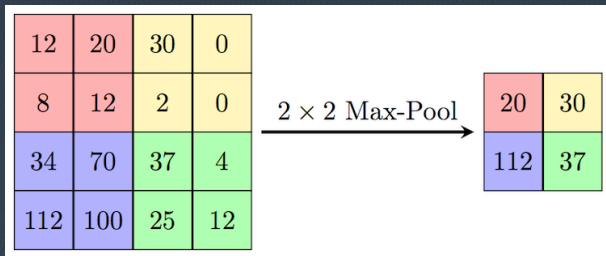
- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
 - CV
- 5 参考文献

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
 - CV
- 5 参考文献

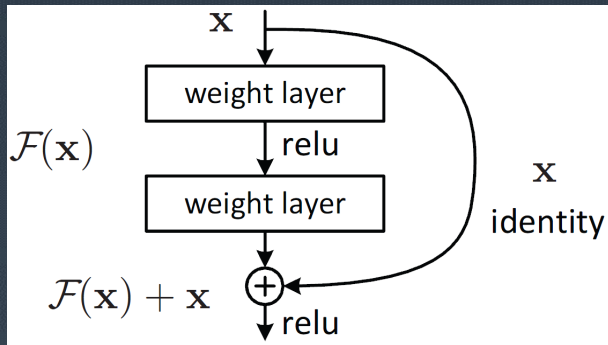




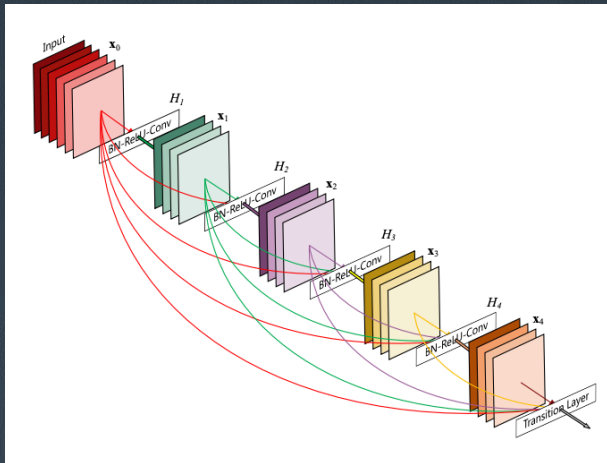
- ▶ stride
- ▶ padding
- ▶ dilation



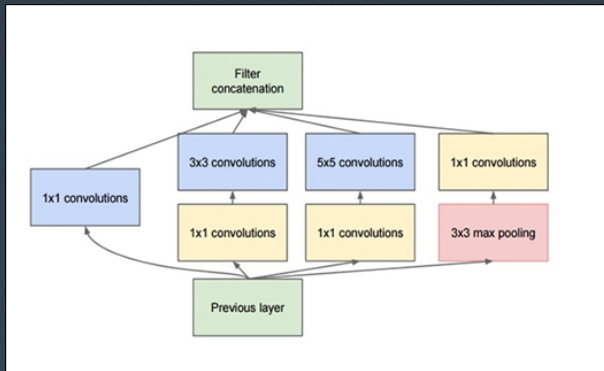
Residual Connection



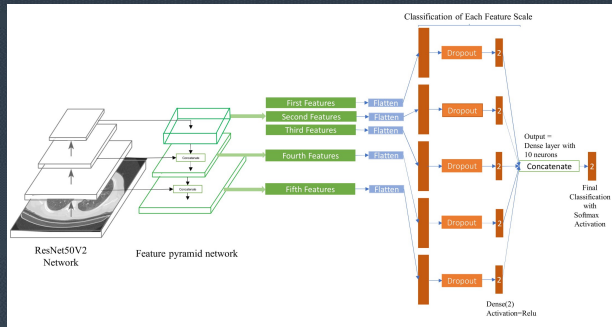
Dense Connection

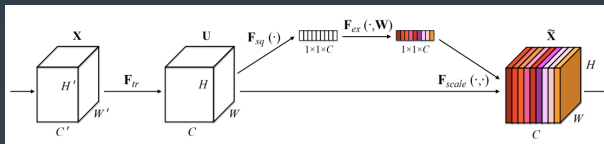


(1) Deep Network	$\mathbf{x}_{i+1} = F(\mathbf{x}_i)$
(2) Residual Network	$\mathbf{x}_{i+1} = \mathbf{x}_i + F(\mathbf{x}_i)$
(3) Deep Network + Norm	$\mathbf{x}_{i+1} = \text{Norm}(F(\mathbf{x}_i))$
(4) Residual Network + Pre-Norm	$\mathbf{x}_{i+1} = \mathbf{x}_i + F(\text{Norm}(\mathbf{x}_i))$
(5) Residual Network + Post-Norm	$\mathbf{x}_{i+1} = \text{Norm}(\mathbf{x}_i + F(\mathbf{x}_i))$
(6) ReZero	$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i F(\mathbf{x}_i)$





Pyramid Network





- ▶ CV 领域是网络设计十分丰富的领域；
- ▶ 基本网络设计可以从 Image Classification, Object Detection 和 Image Segmentation 应用当中寻找。

- 1 大纲
- 2 神经网络优化过程
- 3 神经网络常见元素
- 4 常见神经网络设计
- 5 参考文献

-  He, Kaiming et al. (2015). “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.
-  Summers, Cecilia and Michael J Dinneen (2019). “Four things everyone should know to improve batch normalization”. In: *arXiv preprint arXiv:1906.03548*.