深度学习中的正则化

2018年10月31日

京东

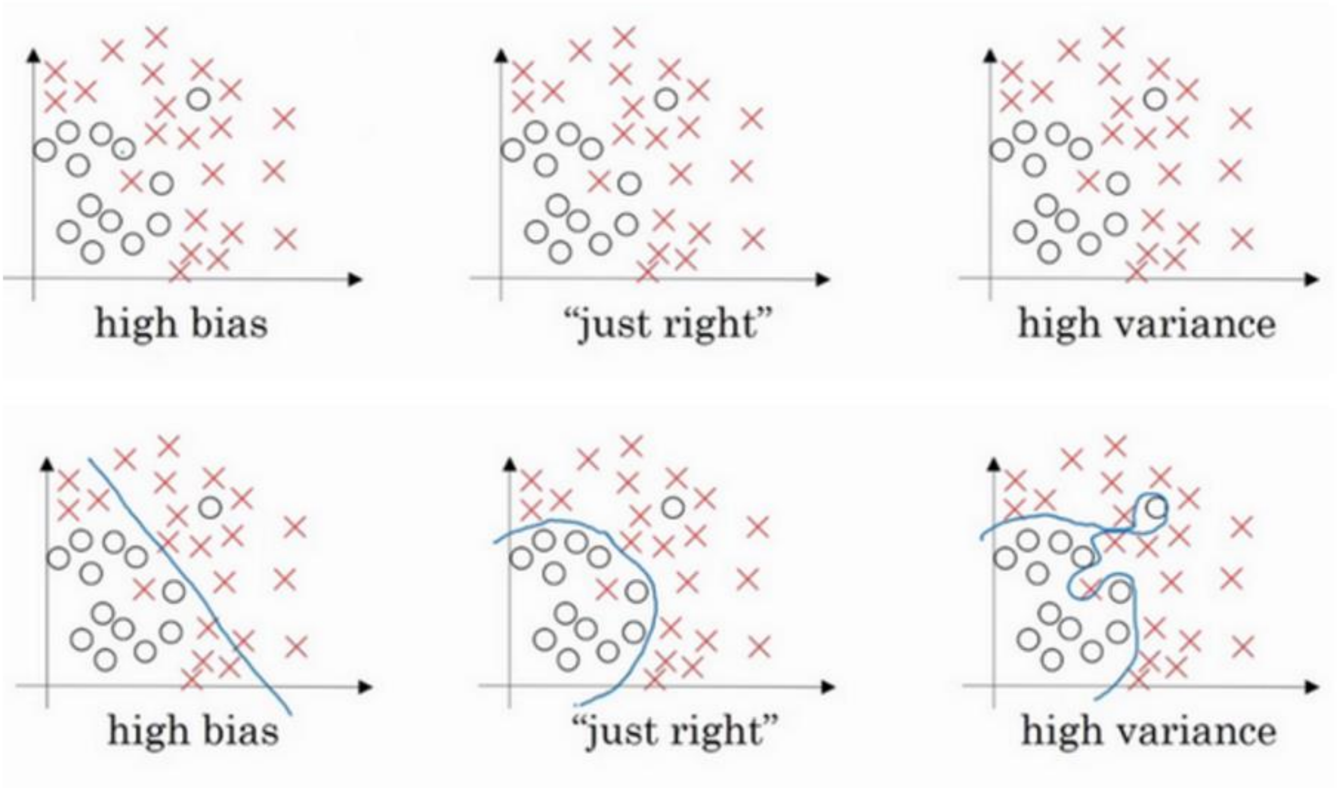
过拟合问题

解决过拟合问题

深度学习中的正则化

过拟合问题

深度学习中的过拟合问题：高方差



解决过拟合问题

1.扩增训练数据
2.正则化

正则化定义

1. 对学习算法的修改-旨在减少泛化误差并非训练误差
2. 传统机器学习的正则化等价于结构风险最小化
3. 深度学习正则化目的与传统机器学习相同

深度学习中的正则化

- 参数范数惩罚
- 数据集增强
- 多任务学习
- Early Stopping
- 参数绑定和参数共享
- 噪声鲁棒性
- 稀疏表示
- 对抗训练

噪声鲁棒性

- 作用于输入—数据增强
- 作用于隐藏单元—Dropout策略
- 作用于权重—主要用于RNN
- 作用于输出标签—标签平滑

标签平滑:

假设k∈{1,2，…，K}是训练数据的预定义类别，其中K是类别的数目。
交叉熵损失表示为：

$$l = -\sum_{k=1}^{K} \log(p(k))q(k)$$

q(x)定义为：

$$q(k) = \begin{cases} 0 & k \neq y \\ 1 & k = y \end{cases}$$

那么交叉熵损失为：

$$l = -\log(p(y))$$

京东

https://arxiv.org/pdf/1701.07717.pdf

$$q_{LSR}(k) = (1-\varepsilon)q(k)+\varepsilon\mu(k),$$

$$其中\mu(k) = \frac{1}{K}$$

$$q_{LSR}(k) = \begin{cases} \frac{\varepsilon}{K} & k \neq y \\ 1 - \varepsilon + \frac{\varepsilon}{K} & k = y \end{cases}$$

$$l_{LSR} = -(1-\varepsilon)\log(p(y)) - \frac{\varepsilon}{K}\sum_{k=1}^{K}\log(p(k)).$$

稀疏表示

- 并不是惩罚模型参数，而是惩罚神经网络中的激活单元，稀疏化激活单元

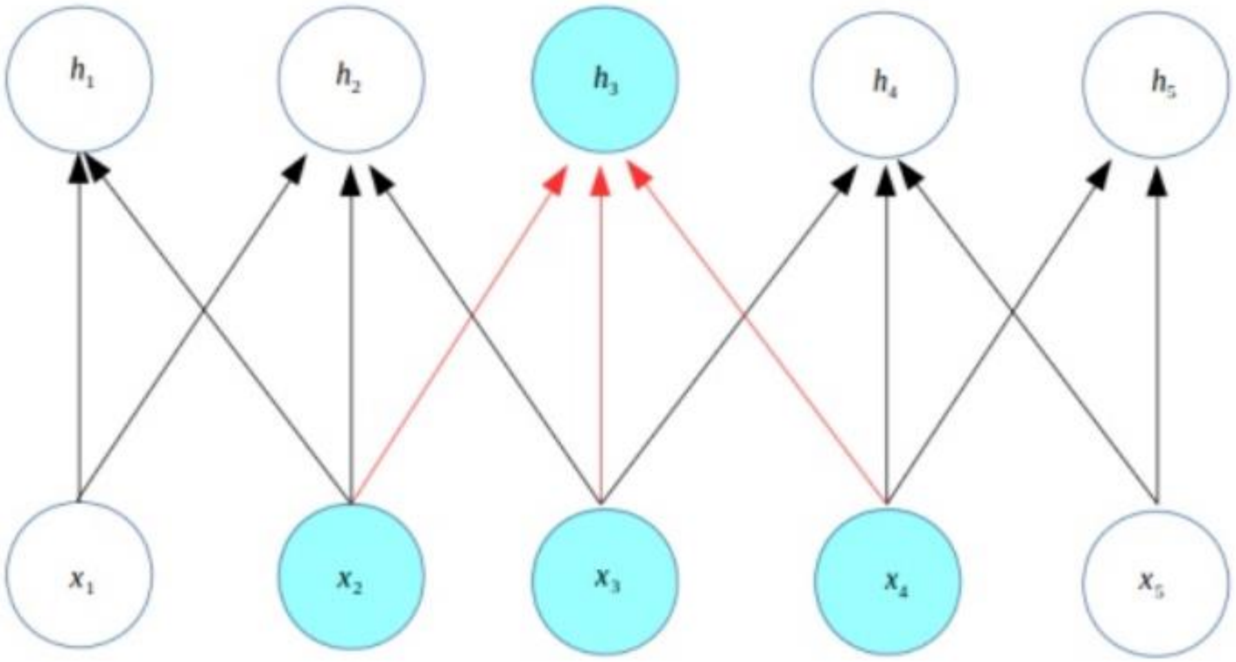
$$\underset{\boldsymbol{y}\in\mathbb{R}^m}{\begin{bmatrix}18\\5\\15\\-9\\-3\end{bmatrix}} = \underset{\boldsymbol{A}\in\mathbb{R}^{m\times n}}{\begin{bmatrix}4&0&0&-2&0&0\\0&0&-1&0&3&0\\0&5&0&0&0&0\\1&0&0&-1&0&-4\\1&0&0&0&-5&0\end{bmatrix}} \underset{\boxed{\boldsymbol{x}\in\mathbb{R}^n}}{\begin{bmatrix}2\\3\\-2\\-5\\1\\4\end{bmatrix}}$$

$$\underset{\boldsymbol{y}\in\mathbb{R}^m}{\begin{bmatrix}-14\\1\\19\\2\\23\end{bmatrix}} = \underset{\boldsymbol{B}\in\mathbb{R}^{m\times n}}{\begin{bmatrix}3&-1&2&-5&4&1\\4&2&-3&-1&1&3\\-1&5&4&2&-3&-2\\3&1&2&-3&0&-3\\-5&4&-2&2&-5&-1\end{bmatrix}} \underset{\boxed{\boldsymbol{h}\in\mathbb{R}^n}}{\begin{bmatrix}0\\2\\0\\0\\-3\\0\end{bmatrix}}$$

- h是x 的一个函数，在某种意义上表示存在于x 中的信息

京东

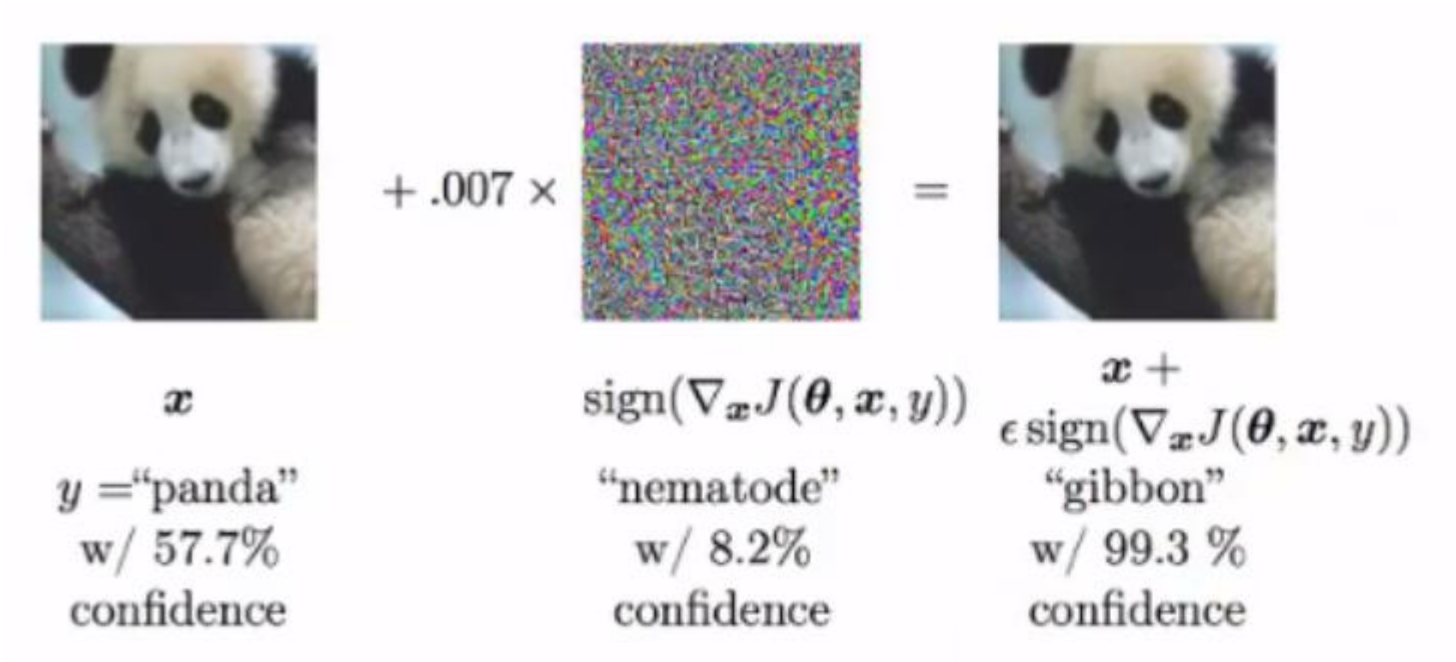
稀疏化激活单元



稀疏表示

对抗训练

目的减少测试集的错误率
对抗样本—在训练集中增加扰动向量
在对抗扰动的训练样本上训练网络



$+ .007 \times$
$=$
$\boldsymbol{x}$
$y =$"panda"
w/ 57.7%
confidence
$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"nematode"
w/ 8.2%
confidence
$\boldsymbol{x} + \epsilon\,\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
w/ 99.3 %
confidence

京东

线性模型中的对抗样本如何产生

$$\hat{x} = x + \eta$$

其中 $\|\eta\|_\infty < \varepsilon$ η为扰动，扰动的每一维都小于ε

对于原ω

$$w^T\hat{x} = w^Tx + w^T\eta$$

$$\eta = sign(\omega)$$

ω有n个维度，每个维度上平均权重为m，那么output可以增加ϵmn

京东

京东