

神经网络

作业五

181220076, 周韧哲, zhourz@smail.nju.edu.cn

2020 年 12 月 29 日

Problem 1

Batch normalization 的输出服从什么样的分布 (指出分布中的具体参数)。

Solution. 服从标准正态分布 $\mathcal{N}(0, 1)$, 均值为 0, 方差为 1。

Problem 2

Batch normalization 为什么归一化后还有放缩 (γ) 和平移 (β)?

Solution. γ 和 β 的引入是为了恢复数据本身的表达能力, 对规范化后的数据进行线性变换, 即 $\hat{x} = \gamma x + \beta$, 就是变换的反操作, 特别地, 当 $\gamma = \sigma, \beta = \mu$ 时, 可以实现等价变换, 并且保留了原始输入特征的分布信息。

Problem3

目前有一批病人的身体数据 (体重变化, 血液指标等) 和他们是否患有肺癌的真实标签, 其中患肺癌的样本只占非常小的比例。数据直接送入一个神经网络中, 求问应该使用什么样的初始化? 数据中不同的特征数值差异过大, 求问如何改进能够让网络更好地学习数据中地分布?

Solution. 可以使用高斯分布初始化。数据中不同的特征数值差异过大可以使用 *Batch Normalization* 把每一层网络的输入数据规范化为均值为 0 方差为 1 的正态分布, 能够缓解特征数值差异过大对网络造成的不良影响从而使得网络能够更好地学习数据分布。

Problem4

对神经网络损失函数添加正则化后形式如下, 这里 V 为损失函数, R 为正则化函数, 给定输入 x_i , 神经网络输出为 $f(x_i)$, y_i 为真实标记。

$$\min_f \sum_{i=1}^n V(f(x_i), y_i) + \lambda R(f)$$

- (1) 请说明参数 λ 的意义。
- (2) 参数 λ 的取值是越大越好还是越小越好, 请结合 playground(<https://playground.tensorflow.org>) 上面的实例说明理由。(注: 在 playground 上, 红框选中的区域用于选择正则化函数的种类以及设置参数 λ 的大小。)

Solution. .

- (1) λ 是正则化系数, 用来调整损失项和正则项之间的系数, 控制着 $R(f)$ 的重要程度。随着 λ 的增大, $R(f)$ 的限制越强, 对模型复杂性的惩罚因子就越大。
- (2) λ 越小, 其惩罚项值不大, 容易出现过拟合现象; λ 越大, 过拟合现象的程度越来越低, 但是当 λ 的值超过一个阈值时, 就会出现欠拟合现象, 因为其惩罚项太大, 导致丢失太多的特征, 甚至一些比较重要的特征。因此, λ 不是越大越好或者越小越好, 它是一个超参数, 需要根据问题实例来调整。如下图 222, 正则化系数分别为 0.001, 0.1, 0.3, 0.3 时完全欠拟合, 网络没有学到任何东西。

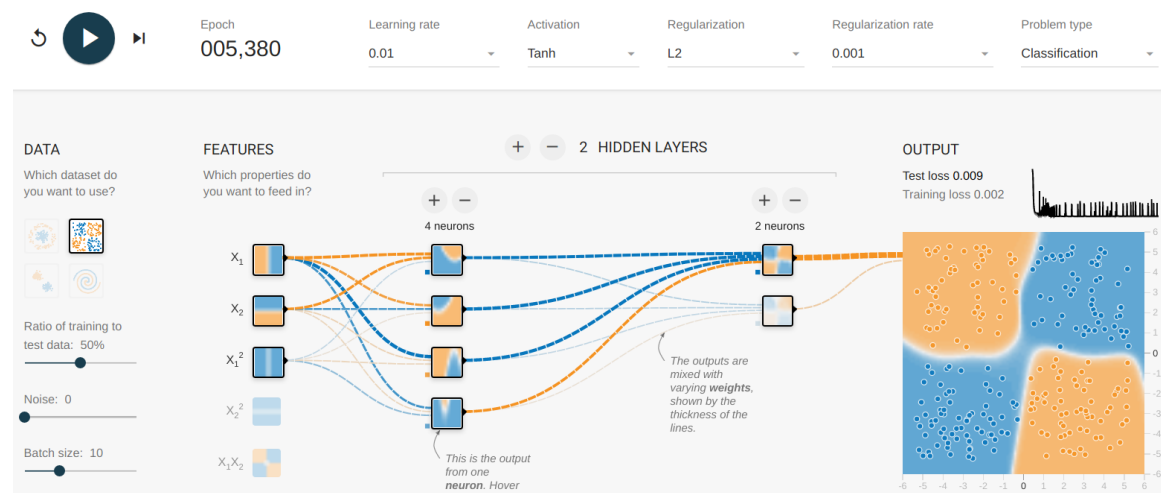


图 1: 0.001

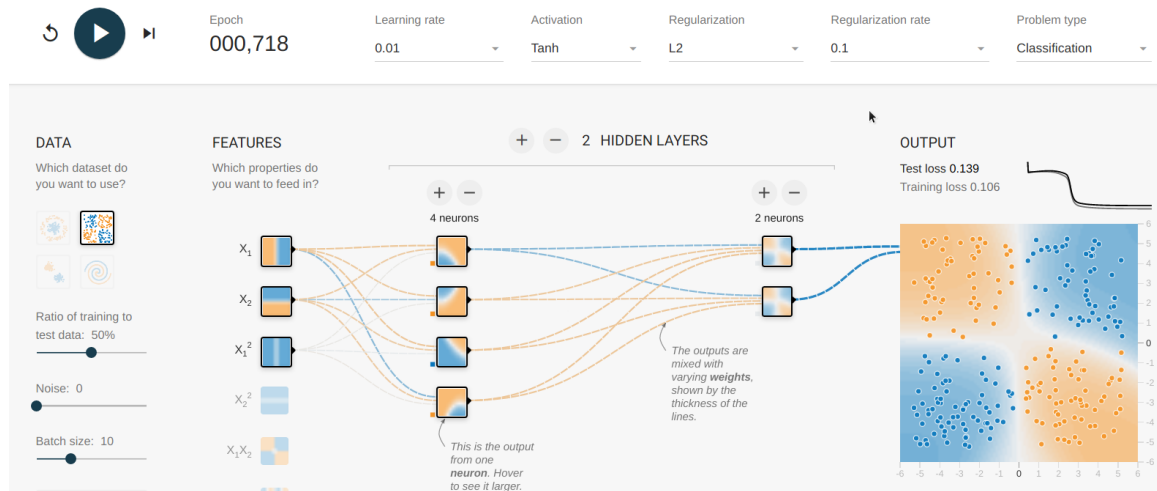


图 2: 0.1

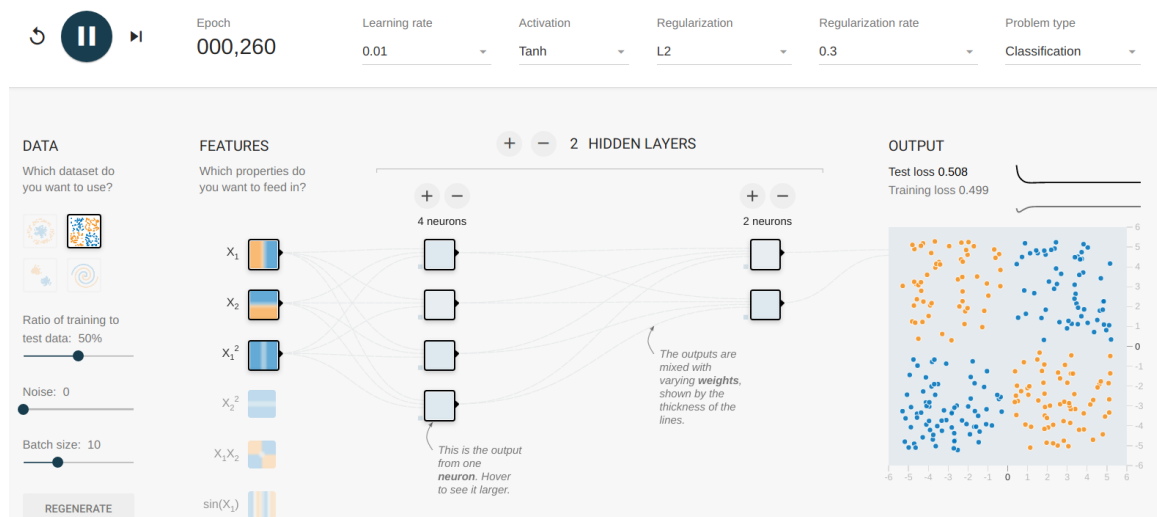


图 3: 0.3