

机器学习Hw2

author: 魏新鹏

student ID: 519021910888

判断题

1. 错误。

没有激活函数，每一层输出的都是上一层输入的线性函数，所以无论网络结构怎么搭，输出都是输入的线性组合，无法非线性分类。激活函数能够增强神经网络的表达能力。

2. 正确。

3. 正确。

4. 错误。

可以对输入进行旋转、平移、缩放等预处理提高模型泛化能力。这就是数据增强（Data Augmentation）

5. 错误。

选择题

1. C

2. C

3. D

简答题

1. 深度学习模型为何在训练中容易出现过拟合？试分析原因并给出如何在深度学习模型训练中缓解过拟合？

如果模型的参数太多，而训练样本又太少，训练出来的模型很容易产生过拟合的现象。

判断方法：模型在验证集合上和训练集合上表现都很好，而在测试集合上表现很差。

可能的原因：

- （1）建模样本选取有误，样本标签错误等，导致选取的样本数据不足以代表预定的分类规则
- （2）样本噪音干扰过大，使得机器将学习了噪音，还认为是特征，从而扰乱了预设的分类规则
- （3）假设的模型无法合理存在，或者说是假设成立的条件实际并不成立
- （4）参数太多，模型复杂度过高
- （5）对于tree-based模型如果我们对于其深度与split没有合理的限制，有可能使节点只包含单纯的事件数据(event)或非事件数据(no event)，使其虽然可以完美匹配（拟合）训练数据，但是无法适应其他数据集
- （6）对于神经网络模型：1.权值学习迭代次数太多(Overtraining)，2. BP算法使权值可能收敛过于复杂的决策面

解决方法：

1. 获取更多的数据，这是解决过拟合最有效的方法，只要给足够多的数据，让模型「看见」尽可能多的「例外情况」，它就会不断修正自己，从而得到更好的结果。比如通过数据增强。
2. 使用合适的模型：通过使用合适复杂度的模型来防止过拟合问题，让其足够拟合真正的规则，同时又不至于拟合太多抽样误差。比如减少网络的层数、神经元个数等均可以限制网络的拟合能力。
3. 限制权值，也叫正则化。

4. 在输入中增加噪声。
 5. 训练多个模型，以每个模型的平均输出作为结果。
 6. Dropout。在训练时，**每次**随机（如50%概率）忽略隐层的某些节点；这样，我们相当于随机从 2^H 个模型中采样选择模型。
2. 深度学习模型训练过程中为何会出现梯度消失和梯度爆炸问题？有哪些方法可以解决梯度消失或梯度爆炸？
(ref: slide:ch11_rnn P23)

本质上是因为神经网络的更新方法，梯度消失是因为反向传播过程中对梯度的求解会产生sigmoid导数和参数的连乘，sigmoid导数的最大值为0.25，权重一般初始都在0，1之间，乘积小于1，多层的话就会有多个小于1的值连乘，导致靠近输入层的梯度几乎为0，得不到更新。梯度爆炸是也是同样的原因，只是如果初始权重大于1，或者更大一些，多个大于1的值连乘，将会很大或溢出，导致梯度更新过大，模型无法收敛。

解决方法：

1. 预训练+微调。

采取无监督逐层训练方法，其基本思想是每次训练一层隐节点，训练时将上一层隐节点的输出作为输入，而本层隐节点的输出作为下一层隐节点的输入，此过程就是逐层“预训练”（pre-training）；在预训练完成后，再对整个网络进行“微调”（fine-tuning）。

2. 梯度剪切+正则化

设置一个梯度剪切阈值，然后更新梯度的时候，如果梯度超过这个阈值，那么就将其强制限制在这个范围之内。这可以防止梯度爆炸。

正则化后的损失函数形式为： $Loss = (y - W^T x)^2 + \alpha ||W||^2$ ，如果发生梯度爆炸，权值的范数会变得非常大，通过正则化可以，限制梯度爆炸的发生。

3. 改变激活函数。

比如使用ReLU，激活函数的导数为1，每层网络都会以相同的速度更新。

4. 使用gated cell来控制哪些信息可以通过。

e.g.: LSTM, GRU, etc