#### 机器学习复习归纳

本人不保证内容完全无误。

Zhc

2018年12月26日，星期三

考试时间：2018.12.27 7:00 PM-9:00 PM

考试地点：A1-302 A1-304

1. 梯度下降法
2. 梯度下降法（Gradient Descent）

原理：每次沿着梯度的反方向移动一小步，最终可以接近一个（局部）最优解。

每次迭代公式：

1. 随机梯度下降法（Stochastic Gradient Descent, SGD）

每次从样本中随机选一个，用来算梯度，迭代。

1. 小批量随机梯度下降法（Mini-batch Stochastic Gradient Descent, MSGD）

随机梯度下降法中每次选一个样本改为固定数量（bench个）的样本。

1. （多元）线性回归
2. 线性回归（Linear Regression）

设有样本m个

拟合模型：

用最小二乘法（Least Square Method）解。

损失函数：，，

损失函数的梯度：

闭式解（解析解）：

1. 逻辑回归（Logistic Regression）

又称对数几率回归。

想将线性回归用在二分类问题上，所以要将线性回归拟合模型的结果映射到区间[0, 1]上，即将原结果再代入Sigmoid函数算一遍。

设，样本有m个，

Sigmoid函数：

拟合模型：

由极大似然法得

损失函数梯度：

1. Softmax回归（Softmax Regression）

用于解决多分类问题。

设有k个分类（），m个样本。

拟合模型：，

或者写成，其中

易得k=2时与逻辑回归相同，可知Softmax回归是逻辑回归的泛化。

损失函数：

（注：课件的损失函数正则化项用的是2范数，但是我看到网络上的资料还有论文用的都是F范数，即矩阵每一项的平方和的算术平方根）

损失函数梯度：

参考：

1. Deep Learning 学习随记（三）Softmax regression <https://www.cnblogs.com/bzjia-blog/p/3366780.html>
2. 李华云. F范数及矩阵分解实例研究[J]. 现代情报, 2008, 28(10):223-225.
3. Softmax回归 <http://deeplearning.stanford.edu/wiki/index.php/Softmax%E5%9B%9E%E5%BD%92>
4. softmax 逻辑回归 <https://baike.baidu.com/item/softmax%20%E9%80%BB%E8%BE%91%E5%9B%9E%E5%BD%92>
5. 范数 <https://baike.baidu.com/item/%E8%8C%83%E6%95%B0>
6. 矩阵范数 <https://baike.baidu.com/item/%E7%9F%A9%E9%98%B5%E8%8C%83%E6%95%B0>
7. 支持向量机（Support Vector Machine）

基本思想：用超平面分隔不同类型的样本，而且样本点离超平面有一定距离。

设超平面为

软间隔（soft margin）：允许部分样本出错。

合叶损失函数（hinge）：

总损失函数：

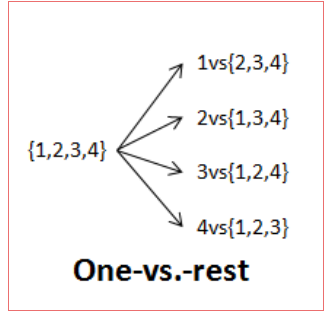
（课件版本：）

1. 用二分类方法解决多分类问题
2. 一对其余（One vs. Rest, OvR）

训练：

对于每个类：

训练一个二分类器，将该内容作为正分类，其余的变成负分类

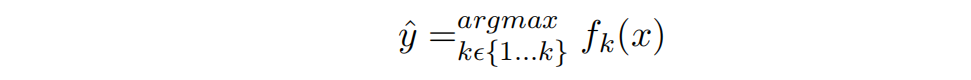


预测：

对于每个二分类器：

产生一个真值信心得分（confidence score）

预测最高得分下的那个label

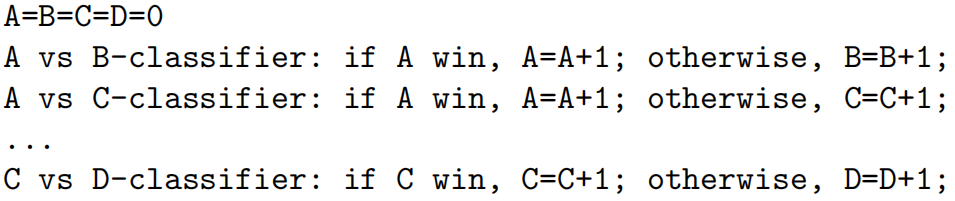


1. 一对一（One vs. One, OvO）

对于每两种类，训练一个二分类器来区分两种类。

有k种就要训练个二分类器。

预测时，对于每个二分类器，比较两个目录内容并且进行投票：



A, B, C, D中谁最大，预测结果就是哪种。

评价：

* 1. 优点：

The distributions of the binary classifications are balanced.

二元分类器的分离是平衡的。

* 1. 缺点：
     1. Train K(K − 1)/2 binary classifiers for a K-way multi-class

problem, which has high computed complexity.

对k路分类问题，训练二分类器，这个能有高计算复杂度。

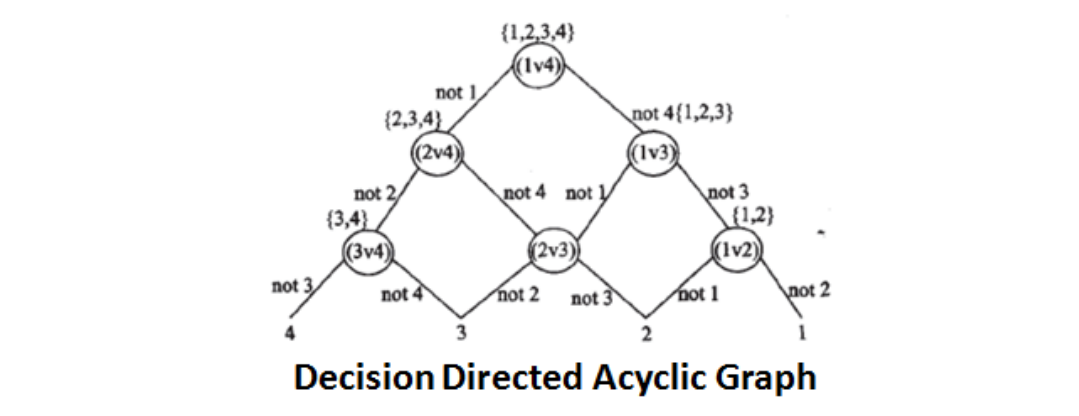
* + 1. Suffer from ambiguities when receive the same number of

votes.

当得到相同的投票的时候，需要忍受模棱两可的选择。

1. 决策有向无环图（Decision Directed Acyclic Graph, DDAG）

比起来一对一方法，此方法使用了一个根基础的二分有向无环图，这个有着内节点和叶子节点：



预测：

从根节点开始

在到达叶子节点之前：

评估二元决策函数

根据输出值向左或者向右移动

比起一对一方法：

每个二元分类器的相关性减少了预测的难度和代价。

1. 其他：神经网络、决策树、k近邻（kNN）
2. 模型评估
3. 过拟合与欠拟合

欠拟合和过拟合的标志：

α 如果通过增加容量我们降低实现错误，那么我们在欠拟合，否则就是过拟合

β 如果测试集展示的错误相当大而且实现错误也是很大，那么就是欠拟合

1. 样本分成训练集、测试集和验证集。
2. 交叉验证法

k折交叉验证（k-fold cross validation）：每k个样本随机留出一个测试集。

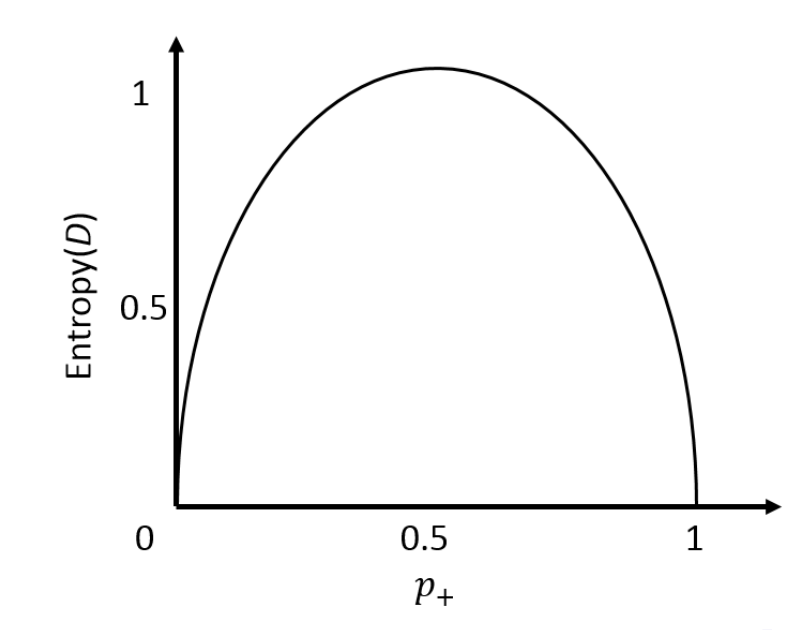
p次重复：将上面的交叉验证重复p次。

被称为：p次k折交叉验证

1. 决策树（Decision Tree）
2. 信息熵（Information Entropy）



其中，样本集合为D，样本有类，为第k类的比例，时约定



1. 信息增益（Information Gain）

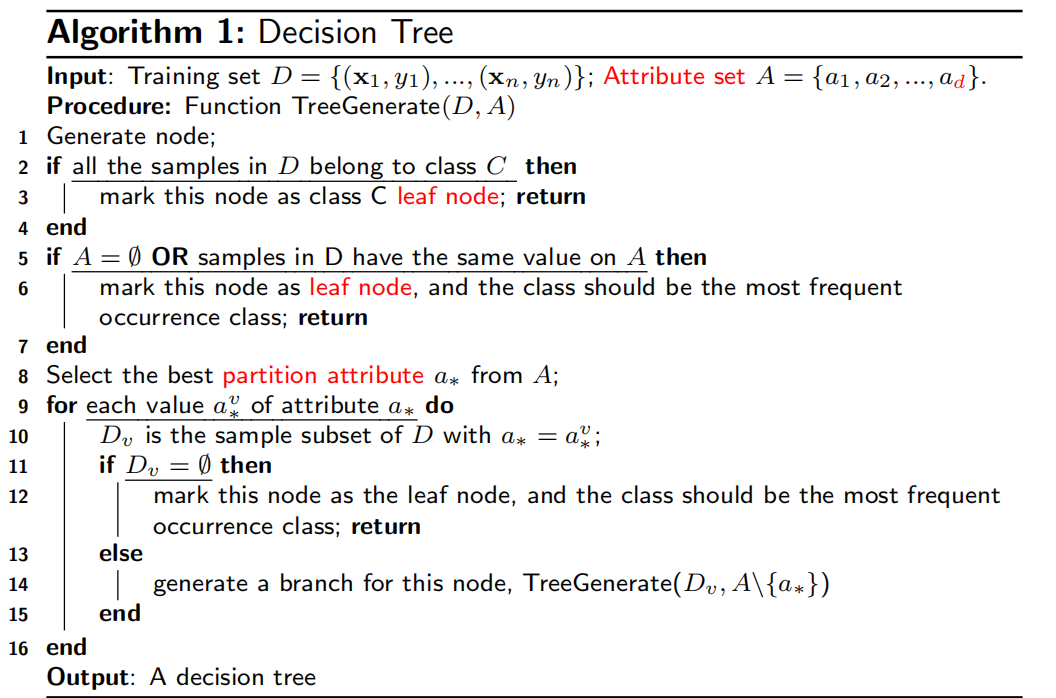


其中表示样本的数量。

1. 决策树

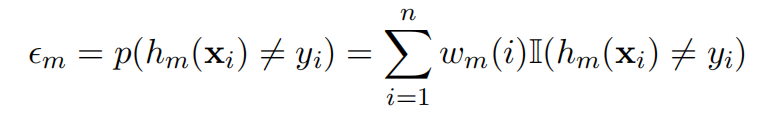
分类分类再分类，直到遇到叶子节点。

训练算法：



1. 集成学习（Ensemble Learning）
2. AdaBoosting

基学习器（Base Learner）：决策树

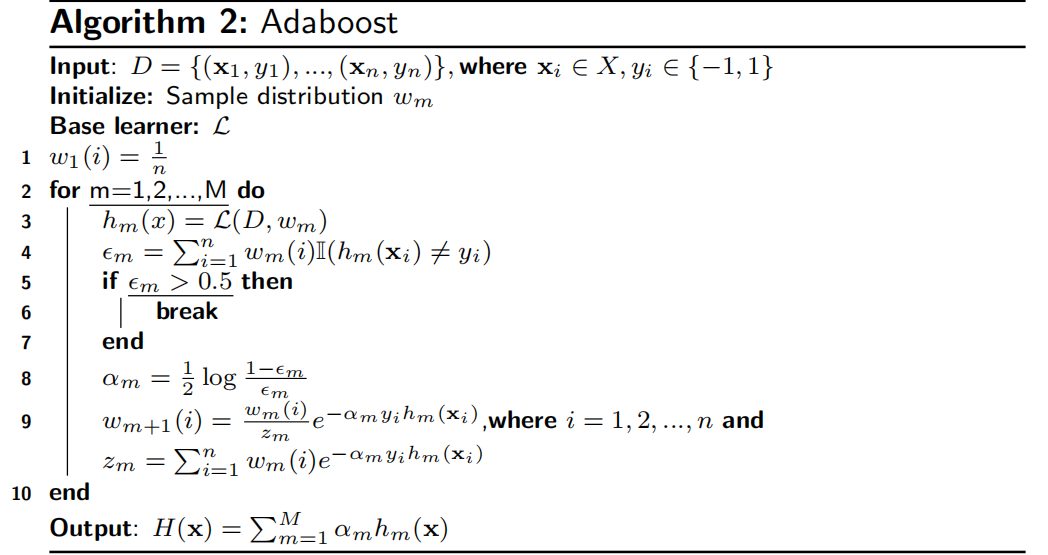
定义错误率：

权重：

拟合模型：，

其中M代表基学习器的数量。

训练算法：



1. 随机森林（Random Forest）：非重点
2. 聚类（Clustering）
3. k均值算法（k-means）

设簇为

损失函数（平方误差）：

均值向量：

训练算法：

α 随机选k个样本作为初始均值向量

β 将所有样本归到欧几里得距离最小的均值向量的簇

γ 计算每个簇的新均值向量

δ 重复β和γ直到均值向量均未更新

1. 层次聚类（Hierarchical Clustering）：非重点
2. 主成分分析（Principal Component Analysis, PCA）

将高维样本通过线性变换映射到低维。

变换公式：，其中是d行d’列的变换矩阵，样本矩阵有m个样本。

主要思路：将所有样本投影到一个超平面上，从而将样本集降维。这个超平面应该具有最近重构性和最大可分性。

最近重构性：样本点到超平面的距离最近。

最大可分性：样本点在超平面上的投影尽可能分开。

由以上两种性质分别可以推导出PCA的损失函数，而且，这两种推导的结论是等价的。

损失函数：，其中d阶方阵，

由拉格朗日乘子法可知，对协方差矩阵进行特征值分解，取最大的个特征向量即可。

算法：

对样本进行中心化：令每个样本向量，其中均值向量

对进行特征值分解

取最大的个特征值所对应的特征值向量得投影矩阵（变换矩阵）

1. 推荐系统（Recommender System）
2. 基于记忆的
3. 基于模型的

核心假设：打分矩阵是近似低秩的，也就是说一个m\*n阶（m行n列）的打分矩阵可以用两个小矩阵和的乘积来近似，即。

损失函数：，其中Frobenius范数（简称F范数）是矩阵所有元素的平方和的算术平方根。

下降方法：交替最小二乘法（Alternating Least Squares）或梯度下降法

交替最小二乘法：一开始随机生成Q，然后固定Q用最小二乘法的解析解算出P，然后固定P用最小二乘法的解析解算出Q，不断交替重复以上两步，直到满足相应条件（如损失函数小于定值或循环次数达到定值）。

每次循环的公式：

设为单位矩阵。

（为与下一条公式比较，便于记忆，可以写成）



为便于记忆，设，，重写以上公式：

模型：

损失函数：

循环公式：





参考：

1. ALS交替最小二乘法 <https://www.jianshu.com/p/f840a5ad1655>
2. Matrix norm <https://en.wikipedia.org/wiki/Matrix_norm>
3. 图像处理
4. 卷积（Convolution）

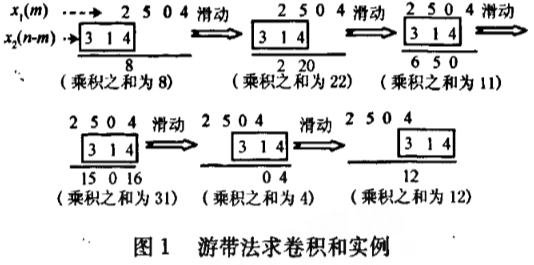
连续函数的卷积：

离散卷积的定义：

有限长序列的卷积：

游带法求卷积：（序列和）

游带法中的滑动实际上就是n递增的过程。注意序列是被翻转过的。



参考：

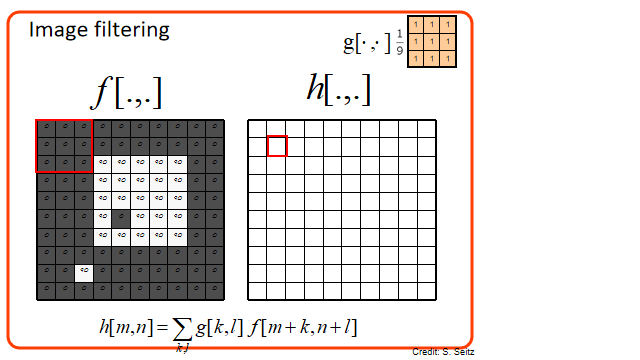
1. 卷积 <https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF>
2. 离散卷积 <https://baike.baidu.com/item/%E7%A6%BB%E6%95%A3%E5%8D%B7%E7%A7%AF>
3. 柯西乘积 <https://baike.baidu.com/item/%E6%9F%AF%E8%A5%BF%E4%B9%98%E7%A7%AF>
4. 李昌利. 有限长序列卷积和求解法[J]. 电气电子教学学报, 2008, 30(1):45-47.
5. 矩阵卷积

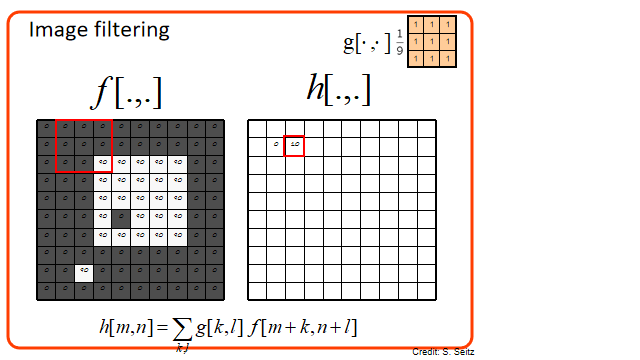
请参考《矩阵卷积、矩阵相乘的转化》 <https://blog.csdn.net/anan1205/article/details/12313593>

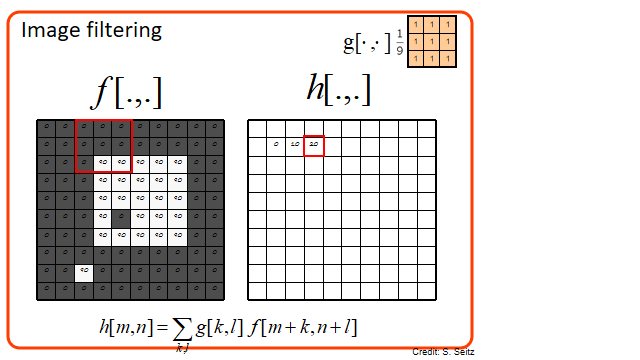
1. 点阵图

由像素矩阵组成，每一个像素有色度、亮度等属性。一幅图片可以看做多个属性矩阵的组合。

1. 用卷积进行图像处理

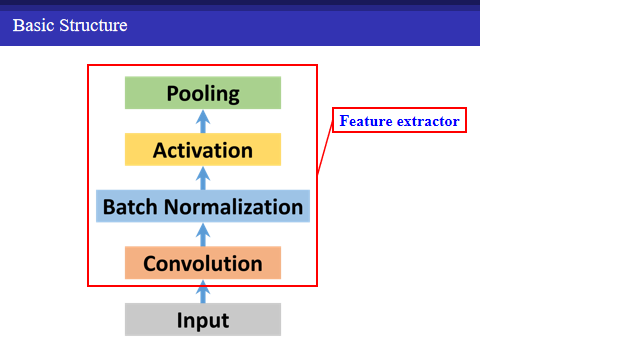






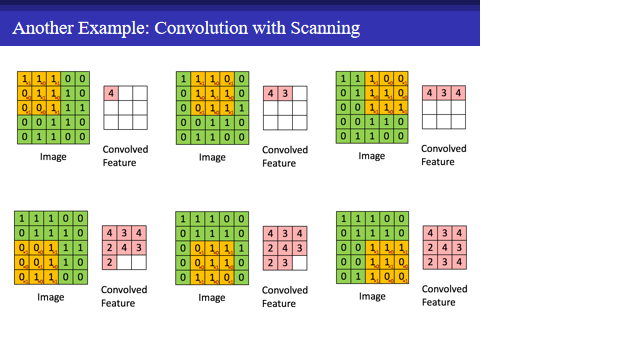
如上。

1. 卷积神经网络
2. 层次概念



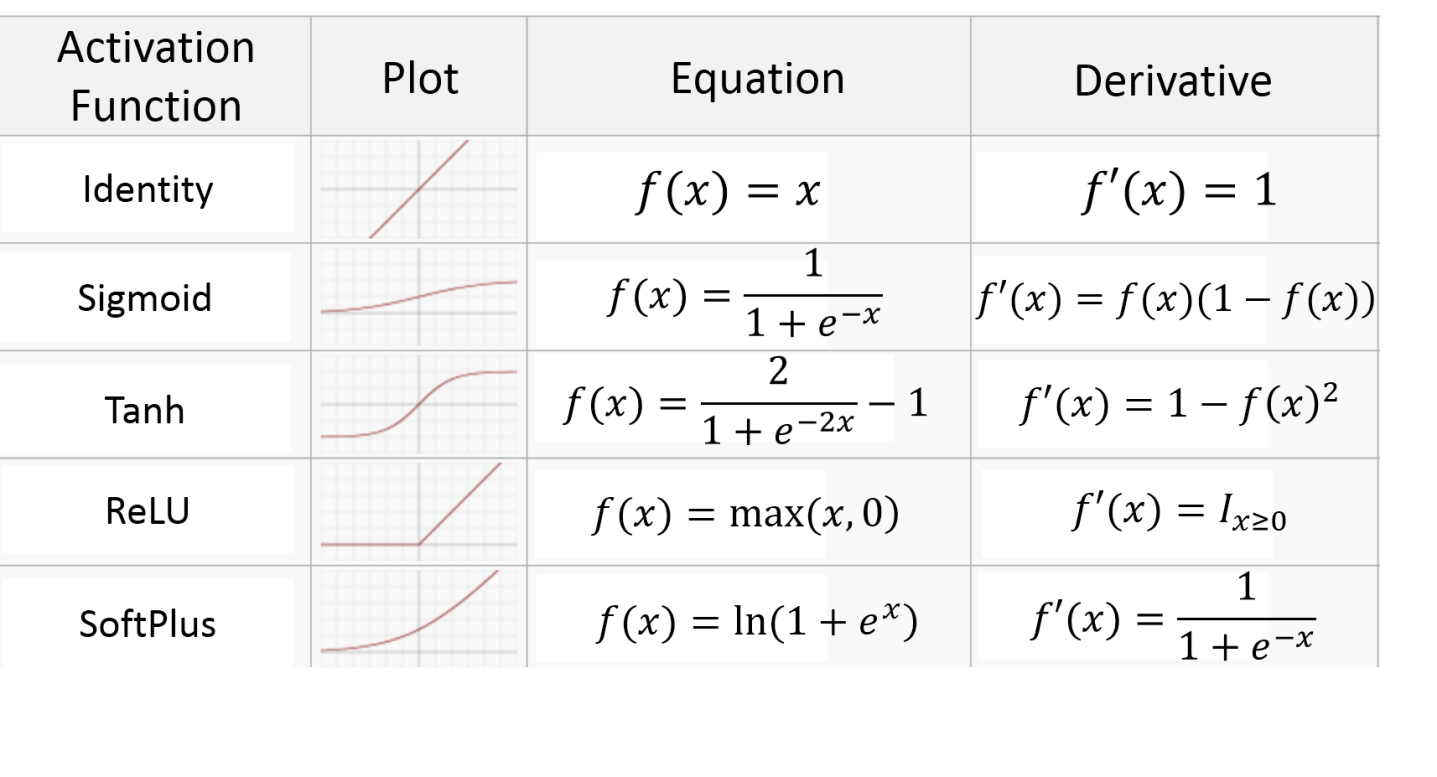
只考卷积层和激活层。

1. 卷积层（Convolution Layer）



1. 激活层（Activation Layer）

选择一个激活函数来计算。



致谢

除了上面标注出来的参考外，本文还参考了何博韬考试内容笔记、赵康铭的《机器学习复习笔记》和周志华的《机器学习》（课本）。