CS229 Lecture notes

原作者：[Andrew Ng](http://cs229.stanford.edu/)（[吴恩达](http://open.163.com/movie/2008/1/M/C/M6SGF6VB4_M6SGHFBMC.html)）

翻译：[CycleUser](https://zhuanlan.zhihu.com/python-kivy)

# 1 感知器（perceptron）和大型边界分类器（large margin classifiers）

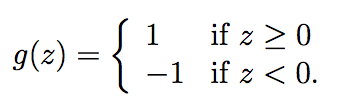
本章是讲义中关于学习理论的最后一部分，我们来介绍另外机器学习模式。在之前的内容中，我们考虑的都是批量学习的情况，即给了我们训练样本集合用于学习，然后用学习得到的假设 h 来评估和判别测试数据。在本章，我们要讲一种新的机器学习模式：在线学习，这种情况下，我们的学习算法要在进行学习的同时给出预测。

学习算法会获得一个样本序列，其中内容为有次序的学习样本，(x(1),y(1)), (x(2),y(2)), ...(x(m),y(m))。最开始获得的就是 x(1)，然后需要预测 y(1)。在完成了这个预测之后，再把 y(1) 的真实值告诉给算法（然后算法就利用这个信息来进行某种学习了）。接下来给算法提供 x(2)，再让算法对 y(2) 进行预测，然后再把 y(2) 的真实值告诉给算法，这样算法就又能学习到一些信息了。这样的过程一直持续到最末尾的样本 (x(m),y(m))。在这种在线学习的背景下，我们关心的是算法在此过程中出错的总次数。因此，这适合需要一边学习一边给出预测的应用情景。

接下来，我们将对感知器学习算法（perceptron algorithm）的在线学习误差给出一个约束。为了让后续的推导（subsequent derivations）更容易，我们就用正负号来表征分类标签，即设 y =∈ {−1, 1}。

回忆一下感知器算法（在第二章中有讲到），其参数 θ ∈ Rn+1，该算法据下面的方程来给出预测：

其中：



然后，给定一个训练样本 (x, y)，感知器学习规则（perceptron learning rule）就按照如下所示来进行更新。如果 hθ(x) = y，那么不改变参数。若二者相等关系不成立，则进行更新1。



当感知器算法作为在线学习算法运行的时候，每次对样本给出错误判断的时候，则更新参数，下面的定理给出了这种情况下的在线学习误差的约束边界。要注意，下面的错误次数的约束边界与整个序列中样本的个数 m **不具有**特定的依赖关系（explicit dependence），和输入特征的维度 n 也**无关**。

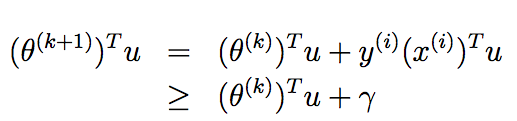
定理 (Block, 1962, and Novikoff, 1962)。设有一个样本序列：(x(1), y(1)), (x(2), y(2)), . . . (x(m), y(m))。假设对于所有的 i ，都有 ||x(i)|| ≤ D，更进一步存在一个单位长度向量 u (||u||2 = 1) 对序列中的所有样本都满足 y(i) · (uT x(i)) ≥ γ（例如，uT x(i) ≥ γ if y(i) = 1, 而 uTx(i) ≤ −γ，若 y(i) = −1，则 u 就以一个宽度至少为 γ 的边界分开了样本数据。）而此感知器算法针对这个序列给出错误预测的综述的上限为 (D/γ)2 。

证明。感知器算法每次只针对出错的样本进行权重更新。设 θ(k) 为犯了第 k 个错误（k-th mistake）的时候的权重。则 θ(1) = ⃗0（因为初始权重为零），若第 k 个错误发生在样本 (x(i), y(i))，则g((x(i))T θ(k)) ≠ y(i) ，也就意味着：



另外根据感知器算法的定义，我们知道 θ(k+1) = θ(k) + y(i) x(i)

然后就得到：

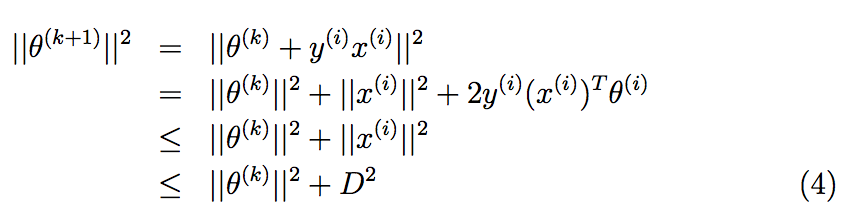


利用一个简单的归纳法（straightforward inductive argument）得到：



1这和之前我们看到的更新规则（update rule）的写法稍微有一点点不一样，因为这里我们把分类标签（labels）改成了 y ∈ {−1, 1}。另外学习速率参数（learning rate parameter） α 也被省去了。这个速率参数的效果只是使用某些固定的常数来对参数 θ 进行缩放，并不会影响生成器的行为效果。

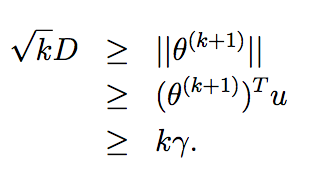


还是根据感知器算法的定义能得到：

上面这个推导过程中，第三步用到了等式(2)。另外这里还要使用一次简单归纳法，上面的不等式(4) 表明：



把上面的等式 (3) 和不等式 (4) 结合起来：



上面第二个不等式是基于 u 是一个单位长度向量（zT u = ||z|| · ||u|| cos φ ≤ ||z|| · ||u||，其中的φ 是向量 z 和向量 u 的夹角）。结果则表明 k ≤ (D/γ)2。因此，如果感知器犯了一个第 k 个错误，则 k ≤ (D/γ)2 。