机器学习基石学习笔记

# 第一课时

介绍本课程的介绍机器学习的结构，

# 第二课时

什么是机器学习？

类比人类的学习过程：

Observation ——learning——skill

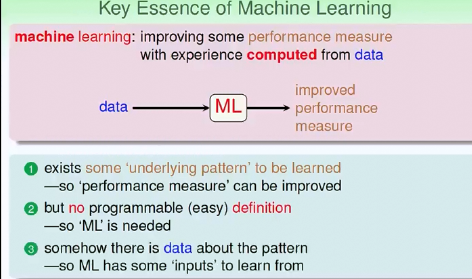
机器学习是：

Data——ML——skill （improved performance/measure）

通过数据分析及训练，得到模型来处理其他数据

机器学习应用：机器学习具有非常广泛的应用的领域，解决许多人难以处理或难以实现的过程。

机器学习特点：一个情境或者一个问题能否应用机器学习或者是否适用机器学习，主要从以下三个特点来分析，满足下面三个特点才能实现机器学习算法分析。



机器学习关键特征

# 第三课时

从衣食住行四个方面分析机器学习的应用需求

机器学习：data ——ML ——skill

举例教育，通过学生答题的历史数据，来预测学生能否答对某个题。

Data：学生答题历史数据

Skill：预测学生能否答对题

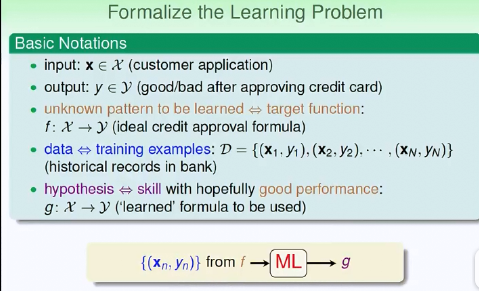
模型：学生答题能力等级 VS 问题难度

通过data 确定学生能力等级与问题难度等级的对应关系，然后针对每一个学生首先评估出其能力，对每个问题得出其难度等级，二者对比预测学生能否答对题。

# 第四课时

资料、data —— ML—— skill

机器学习的关键元素，在建立机器学习模型时要注意分析这5个组成元素：



机器学习组成元素

机器学习学习的目标就是期望找出假设集中的一个假设g尽可能的近似理想的模型f，这里的g 是概率上近似f。

Unkown target function

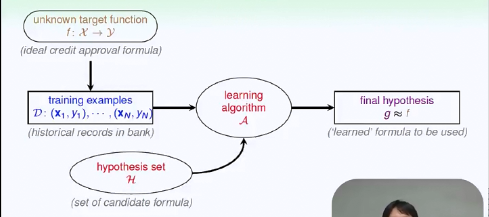
f：x->y

Dataset (x,y)

Learning Algorithm

Hypothesis Set

Final hypothesis



机器学习各组成元素的逻辑关系

# 第五课时

机器学习概念辨析（ML, DM, AI, Statistics）

ML: use data to compute the hypothesis g that approximate f

DM: use huge data to find property that is interesting

（if property 就相当于 hypothesis）机器学习和数据挖掘是很相似的。

从大量数据中找出一个假设，能够应用于相类似的数据预测、分类

AI: compute something that shows intelligent behavior

机器学习是实现AI的一种手段

Statistics: use data to make inference about an unknown process

机器学习中经常应用统计学的理论知识

1~5课时总结

关键内容：

1、机器学习概念：data——ML——skill

2、ML的特点：三个特点

3、ML的关键组成元素：五个组成元素

4、ML关键组成元素之间的逻辑关系：

# 第六课时

介绍第一种假设集：Perceptron Learning Algorithm (PLA)

PLA：就是根据数据特征，建立关于特征的线性模型，通过线性模型和阈值比较来进行二分类或预测。

由对credit card 发放与否问题引出：

Input: X = (x1,x,2,………)

Output: Y = (-1 ,1)

Hypothesis:

引出感知器模型：

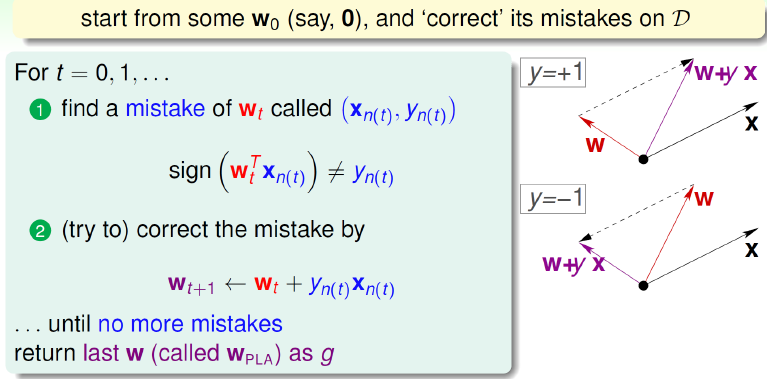
# 第七课时

求解PAL，即从感知器这个假设集中选出一个假设能最好的接近数据集的理想函数模型，也就是介绍PAL的算法逻辑

首先针对线性可分的数据集，才存在一个理想的函数模型，因此首先针对线性可分数据集的PAL解法进行分析。

目标：使假设g 近似理想模型f，但是当f不知道时是很难做到的

解法逻辑：



PAL在线性可分数据集上的解法

Python 代码实现：算法思路的关键就是找出错误点，更正更新w，直到没有错误点。

##############################################################################

#针对线性可分样本集的PAL算法

##############################################################################

def linearPAL(w,Dataset,Classlabels):

from numpy import sign,multiply,array

over = False

#data 样本数

n = len(Dataset)

while(over == False):

Mistakepoints = []

#检测错误点数

for j in range(n):

if sign(sum(multiply(w,Dataset[j,:]))) != classlabels[j]:

Mistakepoints.append(j)

#终止条件

if len(Mistakepoints) == 0:

over =True

#更正w

for i in range(n):

if sign(sum(multiply(w,Dataset[i,:])) != classlabels[i]:

w = array(w) + array(Classlabels[i]\*Dataset[i,:])

return w

#######################################################################################

#针对不可分样本集的PAL算法

#######################################################################################

def findmistakepoint(w,Dataset,Classlabels):

from numpy import sign,multiply

Mistakepoints = []

for i in range(len(Dataset)):

if sign(sum(multiply(w,Dataset[i,:])) != classlabels[i]:

Mistakepoints.append(i)

return Mistakepoints

def correctmistakepoint(w,Mistakepoints,Dataset,Classlabels):

from numpy import array

min\_mistakepoints = Dataset

final\_w = w

for i in range(len(Mistakepoints)):

new\_w = array(w) + array(Classlabels[Mistakepoints[i]]\*Dataset[Mistakepoints[i],:])

new\_mistakepoints = findmistakepoint(new\_w,Dataset,Classlabels)

if len(min\_mistakepoints) > len(new\_mistakepoints):

min\_mistakepoints = new\_mistakepoints

final\_w = new\_w

if len(Mistakepoints) > len(min\_mistakepoints):

w = final\_w

return w

def perceptronAL(w,Dataset,Classlabels,Maxiter):

iter = 0

while(iter < Maxiter):

#定义变量

Mistakepoins = []

#找错误点

Mistakepoins = findmistakepoint(w,Dataset,Classlabels)

#判读是否没有错误点，终止程序

if len(Mistakepoints) == 0:

return w

#改错误点

w,Mistakepoins= correctmistakepoint(w,Mistakepoins,Dataset,Classlabels)

# 第八课时

主要对PAL算法的终止条件进行分析

首先针对线性可分的数据集来分析，集存在没有mistake points 的理想函数模型；

是理想函数模型的权值向量，PAL的实际过程是使 more aligned to with

Linear separable Dataset => exist perfect such that

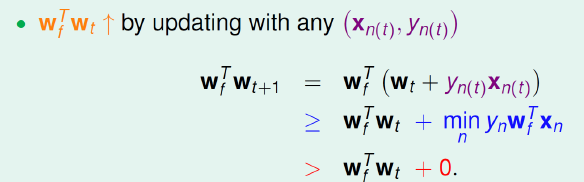
Hypothesis: 在Dataset线性可分的条件下

,其实H表示的是对x的分割面，需要求的也是一个分割面，能分割Dataset 中的点，是分割面的法线向量，因此对Dataset 中的x, H 就是表示点x到分割面的距离，是<,x>的内积。内积有正负，因此乘上标签Y，则均为正。

即, 所以

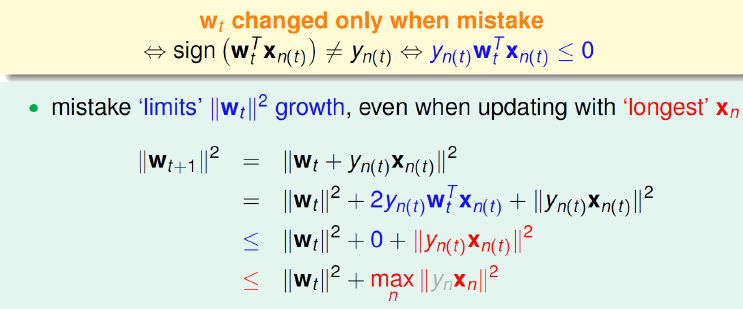
通过Dataset中的点（，）来更新w





为了衡量假设函数与理想模型的接近程度，通过这个内积来衡量，内积越大，表明两个向量的方向的越一致，但是为了消除这两个向量的模长对内积的影响，需要求两个向量的模长来归一化，因此以下是对的模长求解。

直接使用Dataset中的点来更新w，w的模长会增长很慢，



利用对的递归表达式，

可以得到：

=>=

这就表明：由归一化的内积最大为1，内积的最小值与数据集和迭代次数有关。随着T的增大也就也来越接近，同时也表明迭代有上界，最终或在某一个T收敛。

# 第九课时

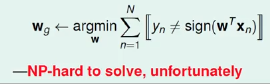
主要介绍如何解决非线性可分数据集的PAL解法：

对非线性的数据集：

assume ‘little’ noise yn = f(xn)

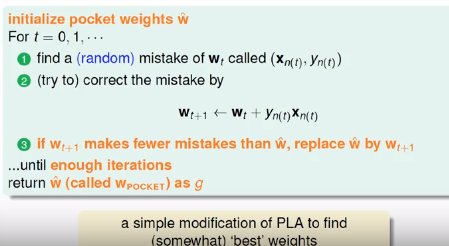
if so, g ~ f， on the Dataset <=> g(xn)

因此这时的目标是使错误点最少，而不是没有错误点了；即



解决办法就是：modify the PAL to get an approximately good g

Modify the PAL algorithm by keeping best weight in pocket.



6~9课时总结

只要介绍了一种假设集类型PAL

PAL 假设集take linear separable Dataset D and perceptron H to get hypothesis g

关键内容：

1、PAL 是什么？ 就是对数据特征进行加权求和，得到的结果与阈值比较来进行分类或者预测。https://images-cdn.shimo.im/uvCoZIxOObMpl7Rg/image.png!thumbnail

2、 PAL 在线性集下的解法： https://images-cdn.shimo.im/gkF4jTM3w5QXGFAx/image.png!thumbnail

3、 PAL在线性集下收敛的证明，即PAL 的中终止条件；https://images-cdn.shimo.im/iyZJbqeb9rg99yvo/image.png!thumbnail

该式表明对线性可分数据集一定是收敛的；

对线性可分数据集，终止条件是： 分类无错误点

对不线性可分的数据集，终止条件是：在一定的迭代次数下，分类错误点数最少

4、PAL在线性不可分数据集下的解法，使用pocket algorithm， 即使用错误点进行更新，但是需要在新的w下的错误点数比原来的w下的错误点数小才更新wl;

# 第十课时

根据输出分类：

1、binary classification ：y ={-1, 1}

2、multiclass classification ： y = {1, 2, 3, ............}

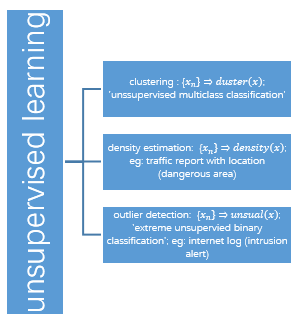
3、structure learning : y = structure

第十一课时

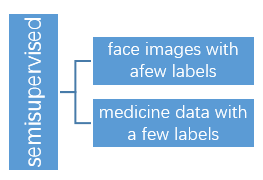
根据是否有标签分类：

1、 supervised: 每一个样本都有一个类别标签

2、unsupervised: 没有标签



3、semisupervised:



4、reinforcement learning ： 通过引入奖惩机制来进行参数修正，从而获得相对满意的输出

# 第十二课时

根据不同的学习策略分类：

1、 batch learning : 批量学习=>应用

2、 online learning ： 边学习，边应用，边改进

3、 active learning ： 这种策略就是要求学习过程中，机器能向人求助提问，机器难以确定的问题，就对人提出请求；这是一种主动学习方式。

online learning ：

1、observe a simple

2、 predict the status with the current https://images-cdn.shimo.im/bqlDZlVVciAEZrln/image.png!thumbnail

3、receive the 'desired label' yt from user, the update the https://images-cdn.shimo.im/SWY4nIeuXzsLkBfS/image.png!thumbnailwith the https://images-cdn.shimo.im/I6xj2PaZR3wU9AzJ/image.png!thumbnail

PLA 应用online learning： PLA 可以很方便的采用online learning 策略：妹进一个样本，都判断是否分类正确，不正确就修正，并与之前的错误数比较，从而决定是否更新w

reinforcement learning应用online learning： 对每一个样本进行处理，利用奖惩机制调整参数，从而逐渐提高模型性能

batch learning : duck feeding

online learning : passive sequensial

active learning : question asking sequential.(询问所选择的的样本xn 的对应标签yn）

# 第十三课时

根据输入的特征抽象程度来分类：

1、 concrete feature

2、 raw feature

3、 abstract feature

分析数据集特征是机器学习的关键所在。

10~13课时总结

主要介绍了机器学习的类

1、根据不同的输出类型分类：分类、回归、结构

Classification、Regression、Structure

2、根据有无标签分类：supervised、unsupervised、semisupervised

Reinforcement learning

3、根据学习策略分类：batch、online、active

4、根据不同输入空间分裂：concrete、raw、abstract

# 第十四课时

分析机器学习的可能性

本课时主要通过一个例子说明：在没有标准的情况，理想模型有很多，假设也有很多，因此并不能找出一个有效的模型。

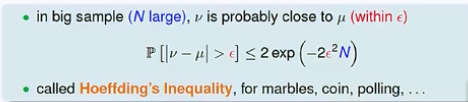
# 第十五课时

假设存在一个理想模型f，使用一个假设g来近似f。

假设模型g对理想模型f，概率上近似，利用hoeffding不等式来证明：

这里以一个箱体中存在红、绿两种颜色的球，这里存在一个红球的真实概率μ，通过从箱体中取样来估计红球的概率，取样得到红球概率为ν。

在大样本下，ν is probably close to μwith ；



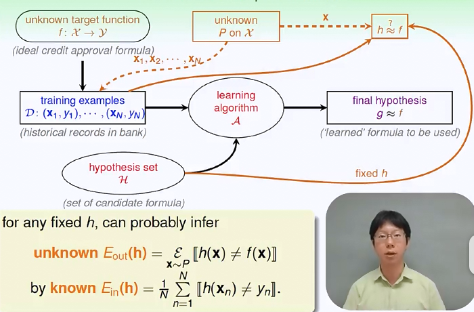
这里的是允许的ν与μ之间的最大误差。

这里分析的是一个样本估计与真实值在概率上的近似

# 第十六课时

将以上的理论应用于机器学习，来分析假设模型与理想模型的概率近似；

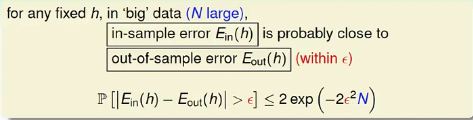
从统计学的角度来分析机器学习；



假设模型的与理想模型的近似是通过各自的错误率是否相近来评估：

:表示假设模型结果与训练集样本标签不同的比例

:表示理想模型结果与测试集样本标签不同的比例（因为理想模型不知道，只能通过假设模型在测试集上的错误比例来代替理想模型的错误比例）



这不等式也说明，存在理想模型时，一定存在一个假设h，使得假设h在概率上近似理想模型f。也就是说假设h对理想模型是：probably approximate correct。

这就将统计概率的理论应用到机器学习可能性的分析之中，假设h是概率上近似f的。

# 第十七课时

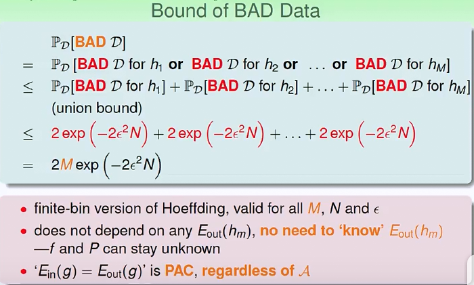
这个课时就是将统计概率理论和真实的机器学习问题想结合：

也就是说真实的情境下，我们并不一定能刚好找到那个能概率上近似理性模型f的假设模型g，真实的情况下是有很多的假设模型。

因此如果在整个假设集上，每个假设的|-|>的概率之和都小于一个设定的常数（总的能接受的假设模型与理想模型不相似的概率），那么对那个能概率上近似理想模型的假设模型，也一定小于设定的常数（也就是能接受的假设模型与理想模型不相似的概率）。

在机器学习中，对一个假设，如果一个样本集训练得到的假设模型h的|-|>，则称这个样本集为BAD sample。

因此结合hoeffding不等式和统计概率理论，能得出以下推导



这就表明：当假设集有限时，只要样本集足够大，总能找到一个概率上近似理想模型的假设模型。

在机器学习可能性分析中，一个关键就是存在一个假设g在概率上近似理想模型。

而这种近似的关键是：样本数N，误差容忍度，只要这两个参数确定，就能分析假设是否存在，机器学习方法是否可能，并不需要和理想模型。

14~17课时总结

主要分析机器学习的可能性：

首先引出机器学习可能找不到一个假设模型来近似理想模型；

然后从抽样统计，引出hoeffding不等式；

再然后将hoeffding不等式引入机器学习。针对单个假设，当样本足够大时，由hoeffding不等式可以证明，假设g是概率上近似理想模型f。

但是真实的情况是，我们并不能刚好找到那个能概率上近似理想模型的假设模型，我们有很多假设模型，我们要保证这很多模型中一定存在概率近似理想模型的假设模型。

注意：我们说概率近似，指的是概率上的不相似度要低于一个人为设定的常数，也是说假设模型和理想模型概率上的相似度要高于某个人为设定的常数。

因此将hoeffding 不等式从单个假设下的概率近似推广到假设集上概率近似。

假设集的大小为M，则对应的hoeffding不等式为：

这里就引出一个问题：

如果M无穷大，那么就会很大，大到人难以接受的程度，也就是难以找到或找不到一个假设模型近似理想模型。

下一课时就会分析在M无限的情况如何进行分析机器学习是否可行。

# 第十八课时

Training VS testing

关键问题：

1、 如何使？

2、如何使

假设集大小M，数据集大小N

M的影响：

1、M is Small，=>很小，也就是和很接近，但是这样可选择的假设很少，概率上近似f的g可能不存在；

2、M is large，=>可能很大，也就是和差别很大，但是可选择的假设很多。

这里关键就是提出假设集大小对两个关键问题的影响，从而进一步分析机器学习的可能性；

# 第十九课时

之前有一个课时说到，当M无限大是，就会很大，从而无法保证能找到假设能概率上近似理想模型。但在本节要说明的问题是：在数据集有限时，假设集也是有限的，并不是无限。因为许多假设非常相似，具有几乎相同模型效果，因此可以归为一类假设，从而使得无限的假设集可以划分为有限类，这就在hoeffding不等式下，保证了很小。即的概率很大

这个不等式，将每个假设看做是独立的，实际上，当时，，，这两个假设具有几乎相同的BAD sample，因此上述表达式，造成了过拟合的问题。解决思路就将相似的假设或具有相同作用的假设归为一类，每个类作为假设集的一个元素。

以PAL算法为例，在确定数目样本的数据集下，分析可能的假设种类，及有多少中划分直线；

PAL算法进行二分类：

1个点：

2个点：

3个点：

4个点：

``````````````````````````

由此可得到假设集大小，将