李航统计学习：

第一章讨论部分：

读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：

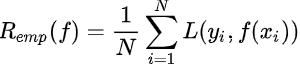
问题1:p17页，为什么根据大数定理，当样本容量趋于无穷时，经验风险趋于期望风险？

讨论后的理解：

期望风险公式如下：



经验风险公式如下：



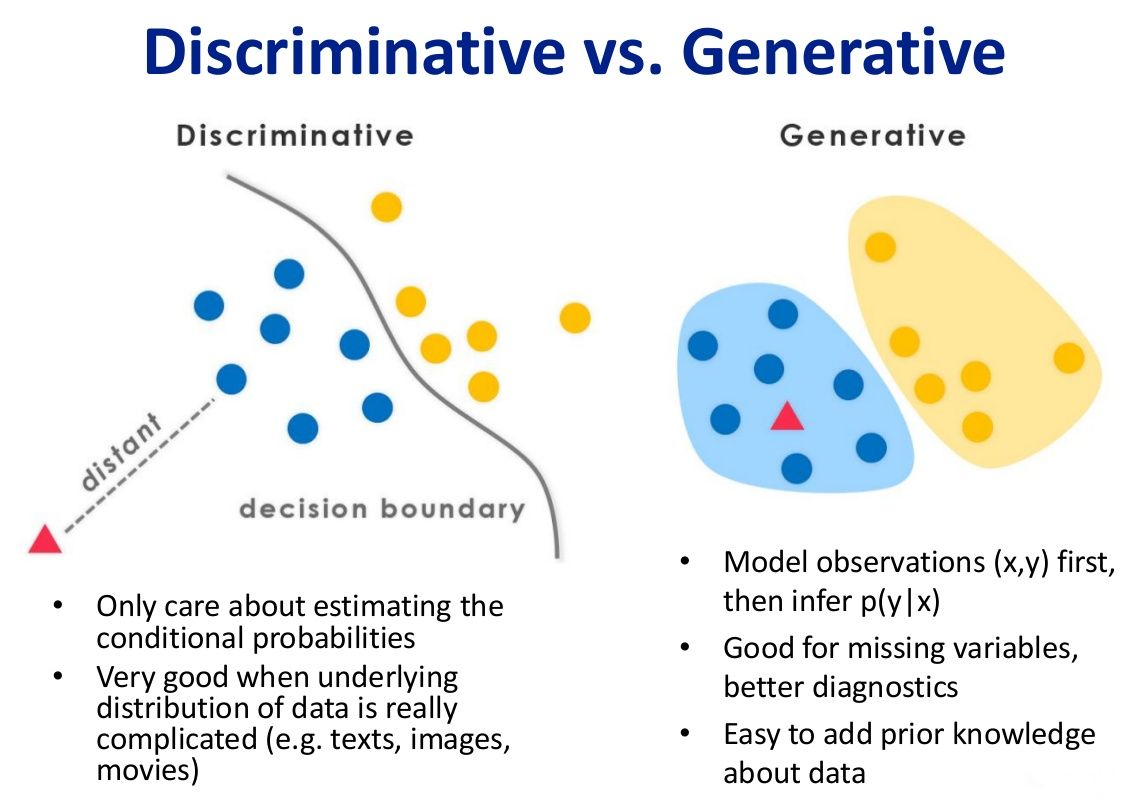
大数定理是指，数量足够大的时候，趋向于无穷，那么样本容量趋向于无穷时，那么样本趋向于总数，频率无限接近于概率。经验风险是根据样本来计算，而期望风险是根据总体数据来计算。那么经验风险必然趋近于期望风险。

1. 提出的问题2：

1.7中，生成方法和判别方法分别适用于什么情况？

讨论后的理解：

在监督学习中，两种方法各有优缺点，适合于不同条件的学习问题。适用的情况图如下：



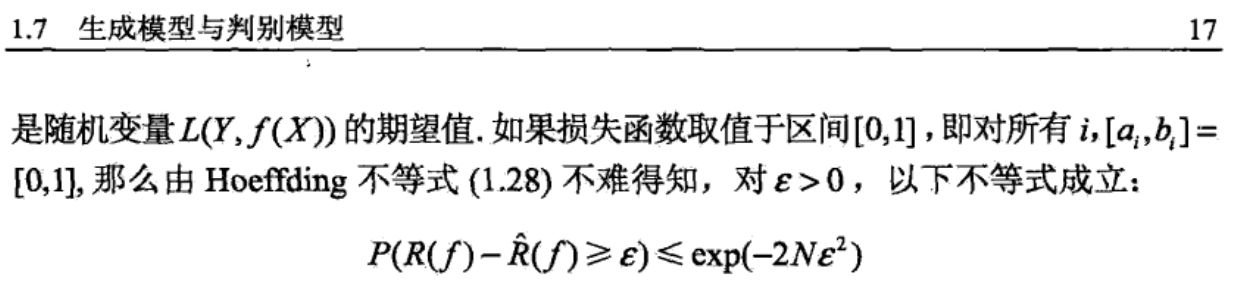
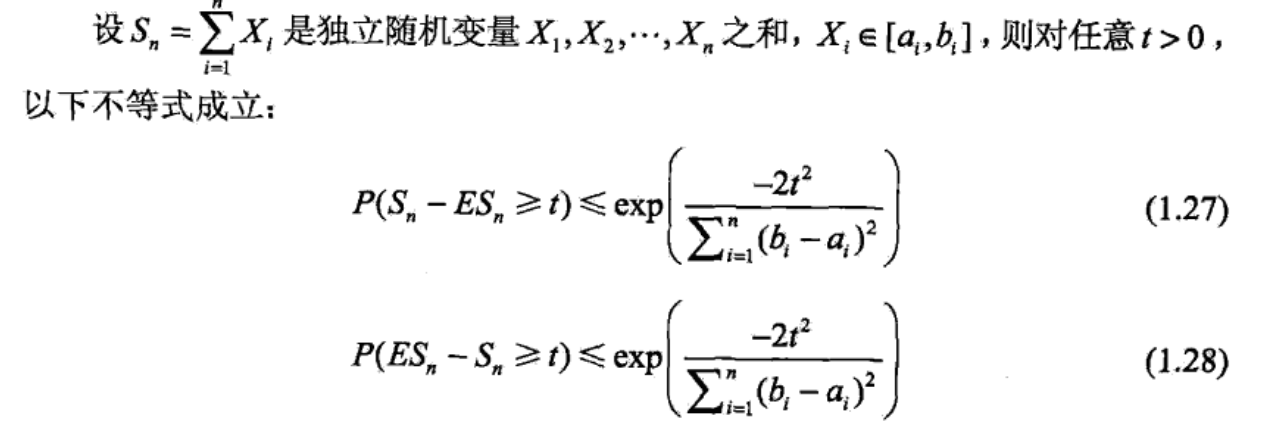
生成方法的特点：上面说到，生成方法学习联合概率密度分布P(X,Y)，所以就可以从统计的角度表示数据的分布情况，能够反映同类数据本身的相似度。但它不关心到底划分各类的那个分类边界在哪。生成方法可以还原出联合概率分布P(Y|X)，而判别方法不能。生成方法的学习收敛速度更快，即当样本容量增加的时候，学到的模型可以更快的收敛于真实模型，当存在隐变量时，仍可以用生成方法学习。此时判别方法就不能用。

判别方法的特点：判别方法直接学习的是决策函数Y=f(X)或者条件概率分布P(Y|X)。不能反映训练数据本身的特性。但它寻找不同类别之间的最优分类面，反映的是异类数据之间的差异。直接面对预测，往往学习的准确率更高。由于直接学习P(Y|X)或P(X)，可以对数据进行各种程度上的抽象、定义特征并使用特征，因此可以简化学习问题。

二、（必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：

问题3：对泛化误差上界的理解？

自己的理解：

泛化误差是检验训练后的模型具不具备代表性，用偏差（bias）和方差（variance）来描述。如果一种方法学习的模型比另外一种方法学习的模型拥有更小的泛化误差上界，那么这种方法就更有效。泛化误差上界是样本容量的函数，当样本容量增加时，泛化误差上界趋于0；泛化误差上界也是假设空间容量的函数，假设空间容量越大，模型越难学，泛化误差上界就越大。

问题4：课后习题1.2,如何证明条件概率分布模型选择对数损失函数就可以通过经验风险最小化推导极大似然估计？

自己的理解：

损失函数是对数的公式为：

L(Y,P(Y|X))=−logP(Y|X)

那么经验风险：

Remp(f)=

即为：

最小化经验风险，那么即为上式的相反数最大化，即为的相反数最大化，即使极大似然估计。

证毕。

问题6：采用结构化风险最小策略时，如何考虑模型的复杂度表示？

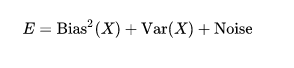
自己的理解：

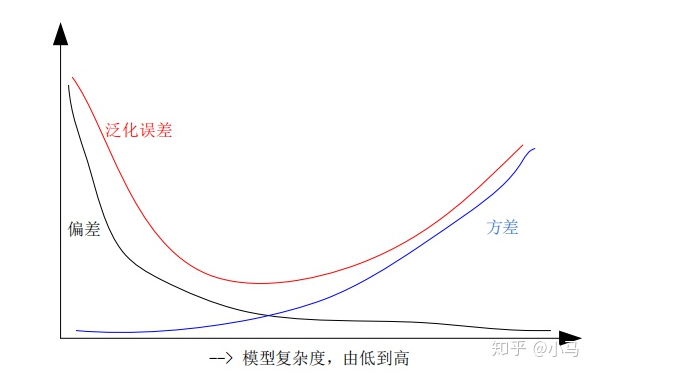
正则化等价于结构风险最小化，其是通过在经验风险项后加上表示模型复杂度的正则化项或惩罚项，达到选择经验风险和模型复杂度都较小的模型目的。

经验风险： 机器学习中的风险是指模型与真实解之间的误差的积累，经验风险是指使用训练出来的模型进行预测或者分类，存在多大的误差，可以简单理解为训练误差，经验风险最小化即为训练误差最小

结构风险： 结构风险定义为经验风险与置信风险（置信是指可信程度）的和，置信风险越大，模型推广能力越差。可以简单认为结构风险是经验风险后面多加了一项表示模型复杂度的函数项，从而可以同时控制模型训练误差和测试误差，结构风险最小化即为在保证模型分类精度(经验风险)的同时，降低模型复杂度，提高泛化能力。

在训练集中训练得到的优秀模型能够很好的适用于实际测试数据，而不仅仅是减少训练误差或者测试误差。泛化误差定义如下：





所以，如上图所示，训练得越多，偏差越小，方差越大，所以要适当的控制模型复杂度，使得泛化误差尽可能的小。

参考知乎链接：https://zhuanlan.zhihu.com/p/97628712

1. （必填）读书计划
2. 本周完成的内容章节：

（1）第一章完成并且梳理知识点，寻找问题，自己思考并且于小组会之前完成了自己的思考。

2、下周计划：第二章的阅读.

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

1. 读书摘要

1.概率模型与非概率模型如何区别？

概率模型就是条件概率建立的模型，选择概率最大的y值作为预测结果。非概率模型就是，使用y＝f(x)的形式，给定一个x值，获得一个y值。

2.损失函数与风险函数的区别。

损失函数是衡量模型预测值f(x)与样本真实值y的误差大小。损失函数是f(x)和y的非负实值函数L。

常见的损失函数有4种。

1).0-1损失函数:如果f(x)与y一致，损失为0，否则为1

2).绝对值损失:L(Y,f(x))＝｜Y-f(x)｜

3).对数损失或对数似然损失:L（Y，P（Y|X））＝－logP（Y|X)（稍微解释一下：Y 是真实值，X是特征，最好的情况时P（Y|X）=1，也就是给定X，模型预测到真实值Y的概率是1，其实也是理想情况下，那么该模型最优，其损失L=-log(1)=0,因为概率取值为0~1,所以需要用-log来衡量损失程度。）

4).平方损失:L(Y,f(x))＝(Y-f(x))²

损失函数值越小，模型越好。

3.风险函数为平均意义下的损失，是期望损失。而损失函数则是一次预测模型好坏的损失。

风险函数＝E(L(Y,f(x)))=∫L(y,f(x))×P(x,y)dxdy

前提假设是x与y是依概率分布P(x,y)，也就是x与y是有关系的，是按照P分布产生的。

经验风险，是关于训练集的平均损失，是损失函数的平均值。

4. 大数定律，在样本量很大的时候，均值接近期望。所以，可以想象，经验损失也就是损失函数的平均值在理想情况下等于期望损失。

但现实数据量远不够，所以之间有一定误差，所以要借用手段进行矫正。其中涉及两种手段，经验风险最小化和结构风险最小化。

5.李航推崇的统计学习方法表达式：

方法=模型+策略+算法

模型，可以指假设空间，也就是你的模型表示范围，模型里有许多未知参数。

策略，是指什么是好模型，即下定义。

算法，是如何按照策略取到指定的最好的参数，类似如何找到最优参数，例如常见的梯度下降，迭代找到最优。由找到的参数也就构建了准确的模型。

1. 训练误差对判断一个问题是否容易学习很重要，但本质不重要，重要的是测试误差，该反映学习方法对未知额测试数据集的预测能力，是学习中的重要概念，也称泛化能力。过拟合是指学习时选择的模型所包含的参数过多，以致于出现该模型对训练集预测很好，但对于测试集效果很差。由下图能看出测试误差与模型复杂度的大体关系。

两种常用的模型选择方法：正则化与交叉验证。

正则化：结构风险最小化策略的实现，是在经验风险加上一个正则化项，一般是模型复杂度的单调递增函数，模型越复杂，正则化值就越大，常见的比如模型参数向量的L2范数。从贝叶斯估计看，正则项对应模型的先验概率，也就指复杂度的模型先验概率较大。

交叉验证：如果样本量很大时，模型选择的一种简单方法是随机地将数据集切分成三部分，分别为训练集、验证集和测试集。训练集用于训练模型，验证集用于模型的选择，而测试集用于最终对学习方法的评估。最终选择在验证集上的最小预测误差的模型。

但是如果没有很多数据，可以采用交叉验证方法。交叉验证的基本想法是重复使用数据；把给定的数据进行切分。

1.简单交叉验证

随机将数据分为训练集和测试集，用训练集在各种参数下进行训练，根据测试集上的测试误差，选出测试误差最小的模型。

2.S折交叉验证

应用最多的方法。首先随机将样本切分S个互不相交的大小相同的子集，然后利用S-1个子集的数据训练模型，利用余下的子集测试样本，将 该过程对可能的S种选择训练集的方法进行训练，取S次评测中平均测试误差最小的模型。

3.留一交叉验证

S折交叉验证的S=N，即只留1个样本作为测试，变换训练集重复N次实验，此处N指样本总数。