5 贝叶斯统计理论

5.1 介绍

- **参数估计：**通过观测到的数据集D，使用概率统计的方法建立模型，这个建立模型的过程就是参数估计的过程。
- **贝叶斯派vs频率派**：两者之间究竟有何不同。
- **MLE和MAP估计**：两者之间的异同点

5.2 后验分布

5.2.1 MAP估计的缺点

1. 没有衡量不确定性：很容易理解，由于是点估计。
2. 使结果过拟合：也是由于是点估计，在预测的时候直接插入使用点估计，过于自信。
3. 众数不是最具代表性的点：(1)在双峰分布中，众数有两个 (2) 斜坡式分布中众数也具代表性。
4. 重新参数化结果改变：重新参数化后，峰值将改变。

5.2.2 置信区间

- 置信区间的定义：包含1-α的概率质量的区间
- 中心区间的定义：中心区间指两边的概率质量相等的区间
- 最高后验密度区间：有许多满足1-α的置信区间，取概率密度最高的那个，也就是最窄的那个。

5.2.3 不同比例的推论

- 举个例子
- 当比较两个观测到的样本总量差异较大时，就可以使用这种方法来进行比较。

5.3 模型选择

1. 交叉验证:也叫CV，常用的有10折交叉验证。
2. 计算模型的后验：注意，是模型的后验，不是参数的后验，这时候根据贝叶斯公式可以知道，需要计算边缘似然。

5.3.1 贝叶斯奥卡姆剃刀

- 奥卡姆剃刀：奥卡姆剃刀的意思并不是说更简单的就更对，而是在两者的表现一样时，才选择更简单的。
- 贝叶斯奥卡姆剃刀：贝叶斯奥卡姆剃刀的意思也是相近的，只是换一个表达方式，假设p(D)越大，模型越好的话。现在有参数更多，更复杂的模型，但它的p(D)不一定越大，就不一定是越好。
- 理解p(D):过拟合时，边缘似然会很小。

5.3.2 计算边缘似然

1. 使用共轭先验计算
2. 使用BIC逼近边缘似然

5.3.3 贝叶斯因子

- 贝叶斯因子的定义

5.4 先验

- 想办法最小化先验的影响，当先验师均匀分布时，还是提供了少量信息，因为参数后验的期望和MLE的结果不一样。

5.4.1 不提供信息的先验

1. 霍尔丹先验，是一个不正确的先验
2. Jeffreys先验，假设不提供信息的先验经过重新参数化后还是一样的。

5.4.2 其他先验

1. Robust先验，当知道先验的某些约束时，使用具有厚尾性的先验，这样先验的影响就会更小。
2. 混合共轭先验：将先验加权平均。

5.5 层次贝叶斯

- 层次贝叶斯假设参数θ也具有不确定行，因此假设η为θ的参数，给η套上先验，从而增加模型的复杂度。
- 也就是全贝叶斯。
- 现在只有η有先验

5.6 经验贝叶斯

- 对η使用点估计，而MLE，MAP是对θ使用点估计。

5.7 贝叶斯决策理论

贝叶斯决策理论讨论的重点放在了预测上，而不再是参数估计了。

- 决策过程：这是对损失函数求期望的最小值
- 贝叶斯估计：这时，损失函数的期望就是损失函数与p(y|x)的加权平均。

5.7.1 常见的损失函数

1. 0-1损失
2. 平方误差损失
3. 绝对误差损失

5.7.2 二分类决策问题

- 精确率、精确率、召回率等区别
- ROC曲线
- PR曲线