Machine Learning

课件 2　　13.10.2021

2021-10-13　02　More on probabilities, Bayes nets

一． 演绎性推理与归纳性推理

Deductive vs. plausible reasoning

演绎推理例子: 「所有人都会死，苏格拉底是人，所以苏格拉底会死。」

由一个大前提和一个小前提，推导出结论的推理逻辑

Deductive reasoning: 演绎性推理

if A is true, then B is true 大前提
A is true　小前提

B is true 结论
"modus ponens" 辩证法

归纳推理例子: 「张三赞同我的回答李,赞同我的回答,王五赞同我的回答」

「我的回答收到了赞同」

找到几个事物的共同点，然后通过共同点去概括这几个事物

Overview:

- Assume: if A is true, then B is true.
  - A is true, implies that B is true.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A is false.

A是真的，意味着B是真的。
A是假的，意味着B变得不太可信。
B是真的，意味着A变得更加可信。
B是假的，意味着A是假的。

- Assume: if A is true, then B becomes more plausible.
  - A is true, implies that B becomes more plausible.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A becomes less plausible.

假设：如果A是真的，那么B就会变得更加可信
A是真的，意味着B变得更加可信。
A是假的，意味着B变得不那么可信。
B是真的，意味着A变得更加可信。
B是假的，意味着A变得不那么可信。

How can we formalize plausibility?

将可行性量化考虑

我们需要将机器学习视为概率建模问题。在机器学习中，模型表述了从某个系统中能观察到的所有数据，也就是说模型不仅可以描述所有我们收集到的某种数据，同时它还能描述那些没收集到的同类数据。

二．考克斯和科尔莫戈罗夫的方法

1.Cox’s and Kolmogorov’s approaches

Cox’s axioms (formalizing common sense) 考克斯公理(常识的形式化)

- plausibility of B assuming that A is true is a real number $p(B|A)$
- plausibility $p(B|A)$ complies with common sense
- plausibility $p(B|A)$ is consistent

考克斯定理（警告：证明不是完全严格的）

Cox’s theorem (WARNING: proof is not completely rigorous)

- product rule: $p(A, B|C) = p(A|B, C)p(B|C) = p(B|A, C)p(A|C)$
- sum rule: $p(A|C) + p(\neg A|C) = 1$

假设A是真的 B的可信度是一个实数p(B|A)

可信度p(B|A)符合常识

可信度p(B|A)是一致的

Notes

- the product rule implies Bayes’ rule

1. 可分性和可比性——命题的合理性是一个实数，取决于我们掌握的与该命题相关的信息。
2. 常识 - 合理性应该随着模型中合理性的评估而显着变化。
3. 一致性——如果一个命题的合理性可以通过多种方式推导出来，那么所有的结果都必须相等

1. 某些真理由 $\Pr(A \mid B) = 1$，以及某些虚假 $\Pr(A \mid B) = 0$.
2. $\Pr(A \mid B) + \Pr(\text{not } A \mid B) = 1$.
3. $\Pr(AB \mid C) = \Pr(A \mid C)\Pr(B \mid AC) = \Pr(B \mid C)\Pr(A \mid BC)$.

作为贝叶斯概率正确性验证的理由(没啥用)

2. Kolmogorov’s approaches

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

第一公理（非负性） 对于任意的事件 P(a) >=0

第二公理（归一化） $P(\Omega) = 1$

第三公理（可加性）

从柯尔莫果洛夫公理可以推导出另外一些对计算概率有用的法则。

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$P(\Omega - E) = 1 - P(E),$$

$$P(A \cap B) = P(A) \cdot P(B|A),$$

这一关系给出了贝叶斯定理。以此可以得出A和B是独立的当且仅当

$$P(A \cap B) = P(A) \cdot P(B),$$

Note:
- Debate: is Bayes' rule an axiom?
- Plausibilities are just probabilities.

归纳性只是概率

从两个 axions(公理)推导出 Bayes‘ rule（规则）

三． 概率的规则

Rules of probability

如果当我们对世界的某种数据建模时，我们需要预测那些没观察到的数据以及它们之间的不确定性，因此我们可以使用数学中概率论描述这种不确定性并完成「模型」的构建。

贝叶斯法则可以描述为「执果索因」，即知道某个事件发生了后，求得该事件最可能是在什么情况下发生的。

贝叶斯法则会告诉我们如何更新对未知世界或假设（hypothesis）的知识与信念，且更新假设或信念的信息从我们已知的观察或数据（data）中获取

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

。

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{hypothesis})P(\text{data}|\text{hypothesis})}{\sum_{\text{h}} P(\text{h})P(\text{data}|\text{h})}$$

在贝叶斯法法则中，总体数据是未知的，也就是说我们需要使用概率分布表征这种不确定性。在观察到数据之前，我们有先验概率 P(hypothesis)，它表示我们事先认为任意一个假设及其所出现的概率。此外，对于任意一个假设，我们需要评估根据它能观察到数据的概率，即 P(data|hypothesis)。

- the evidence $p(\text{data})$ can be calculated from known quantities using the sum rule, $p(\text{data}) = p(\text{data}|H)p(H) + p(\text{data}|\neg H)p(\neg H)$

- Bayes' rule calculates your posterior belief $p(H|\text{data})$ about the hypothesis after seeing the data

Bayes' rule tells you how to update your beliefs!

贝叶斯规则计算出你在看到数据后对假设的后验信念 p(H|data)。

假设的后验信念

贝叶斯规则告诉你如何更新你的信念!

Bayes' rule (4) — monster vs. mouse

在 Bayes 里引入联合概率一起求解

Basic laws of probability:

$p(A,B) = p(A\|B)p(B) = p(B\|A)p(A)$	product rule
$p(A,B\|C) = p(A\|B,C)p(B\|C) = p(B\|A,C)p(A\|C)$	product rule
$p(A) + p(\neg A) = 1$	sum rule
$p(A\|C) + p(\neg A\|C) = 1$	sum rule
$p(A,B) + p(A,\neg B) = p(A)$	sum rule

Lemma 2.1

$p(B|A) = 1$ *implies*意味着

- $p(B|A) = 1$ *"modus ponens"*
- $p(B|\neg A) \le p(B)$
- $p(A|B) \ge p(A)$
- $p(\neg A|\neg B) = 1$ *"modus tollens", alternative* $p(A|\neg B) = 0$

Lemma 2.2

$p(B|A) \ge p(B)$ *implies*

- $p(B|A) \ge p(B)$
- $p(B|\neg A) \le p(B)$
- $p(A|B) \ge p(A)$
- $p(\neg A|\neg B) \ge p(\neg A)$