

# 机器学习白板系列

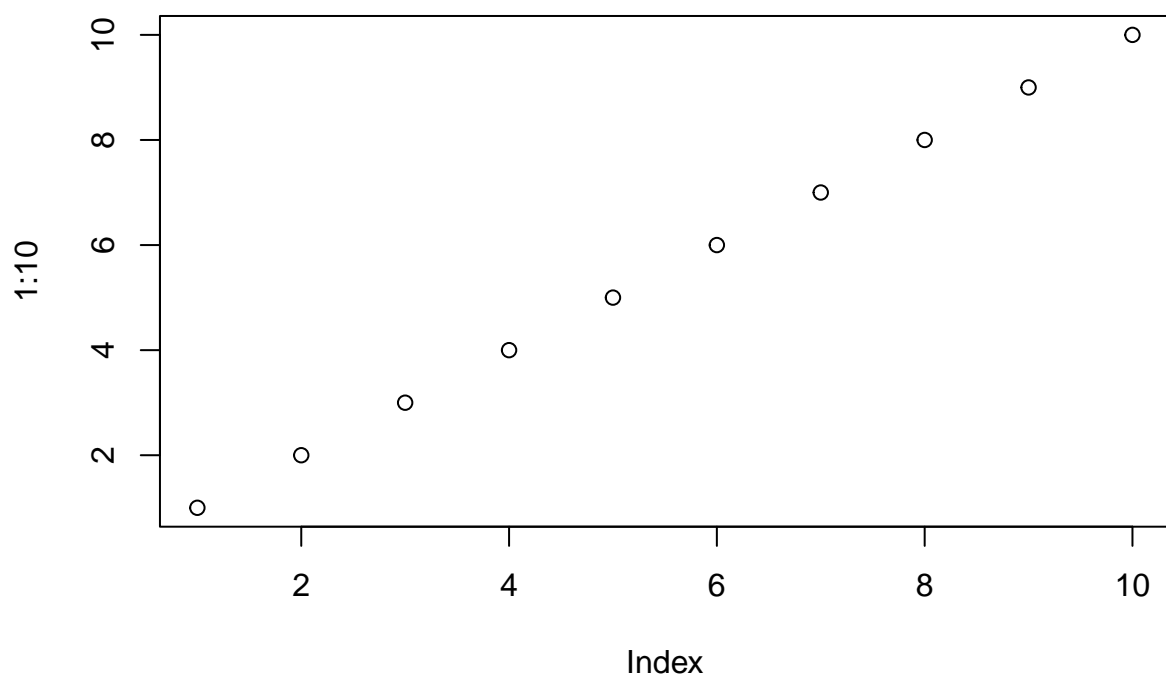
庄闪闪

2022/7/15

# 目录

# 简介

```
plot(1:10)
```



# Chapter 1

## Introduction

对概率的诠释有两大学派，一种是频率派另一种是贝叶斯派。后面我们对观测集采用下面记号：

$$X_{N \times p} = (x_1, x_2, \dots, x_N)^T, x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$$

这个记号表示有  $N$  个样本，每个样本都是  $p$  维向量。其中每个观测都是由  $p(x|\theta)$  生成的。

### 1.1 频率派的观点

$p(x|\theta)$  中的  $\theta$  是一个常量。对于  $N$  个观测来说观测集的概率为  $p(X|\theta) = \prod_{i=1}^N p(x_i|\theta)$ 。为了求  $\theta$  的大小，我们采用最大对数似然 MLE 的方法：

$$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} \log p(X|\theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N \log p(x_i|\theta)$$

### 1.2 贝叶斯派的观点

贝叶斯派认为  $p(x|\theta)$  中的  $\theta$  不是一个常量。这个  $\theta$  满足一个预设的先验的分布  $\theta \sim p(\theta)$ 。于是根据贝叶斯定理依赖观测集参数的后验可以写成：

$$p(\theta|X) = \frac{p(X|\theta) \cdot p(\theta)}{p(X)} = \frac{p(X|\theta) \cdot p(\theta)}{\int_{\theta} p(X|\theta) \cdot p(\theta) d\theta}$$

为了求  $\theta$  的值，我们要最大化这个参数后验 MAP：

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} p(\theta|X) = \underset{\theta}{\operatorname{argmax}} p(X|\theta) \cdot p(\theta)$$

其中第二个等号是由于分母和  $\theta$  没有关系。求解这个  $\theta$  值后计算  $\frac{p(X|\theta) \cdot p(\theta)}{\int_{\theta} p(X|\theta) \cdot p(\theta) d\theta}$ ，就得到了参数的后验概率。其中  $p(X|\theta)$  叫似然，是我们的模型分布。得到了参数的后验分布后，我们可以将这个分布用于预测贝叶斯预测：

$$p(x_{new}|X) = \int_{\theta} p(x_{new}|\theta) \cdot p(\theta|X) d\theta$$

其中积分中的被乘数是模型，乘数是后验分布。

### 1.3 小结

频率派和贝叶斯派分别给出了一系列的机器学习算法。频率派的观点导出了一系列的统计机器学习算法而贝叶斯派导出了概率图理论。在应用频率派的 MLE 方法时最优化理论占有重要地位。而贝叶斯派的算法无论是后验概率的建模还是应用这个后验进行推断时积分占有重要地位。因此采样积分方法如 MCMC 有很多应用。