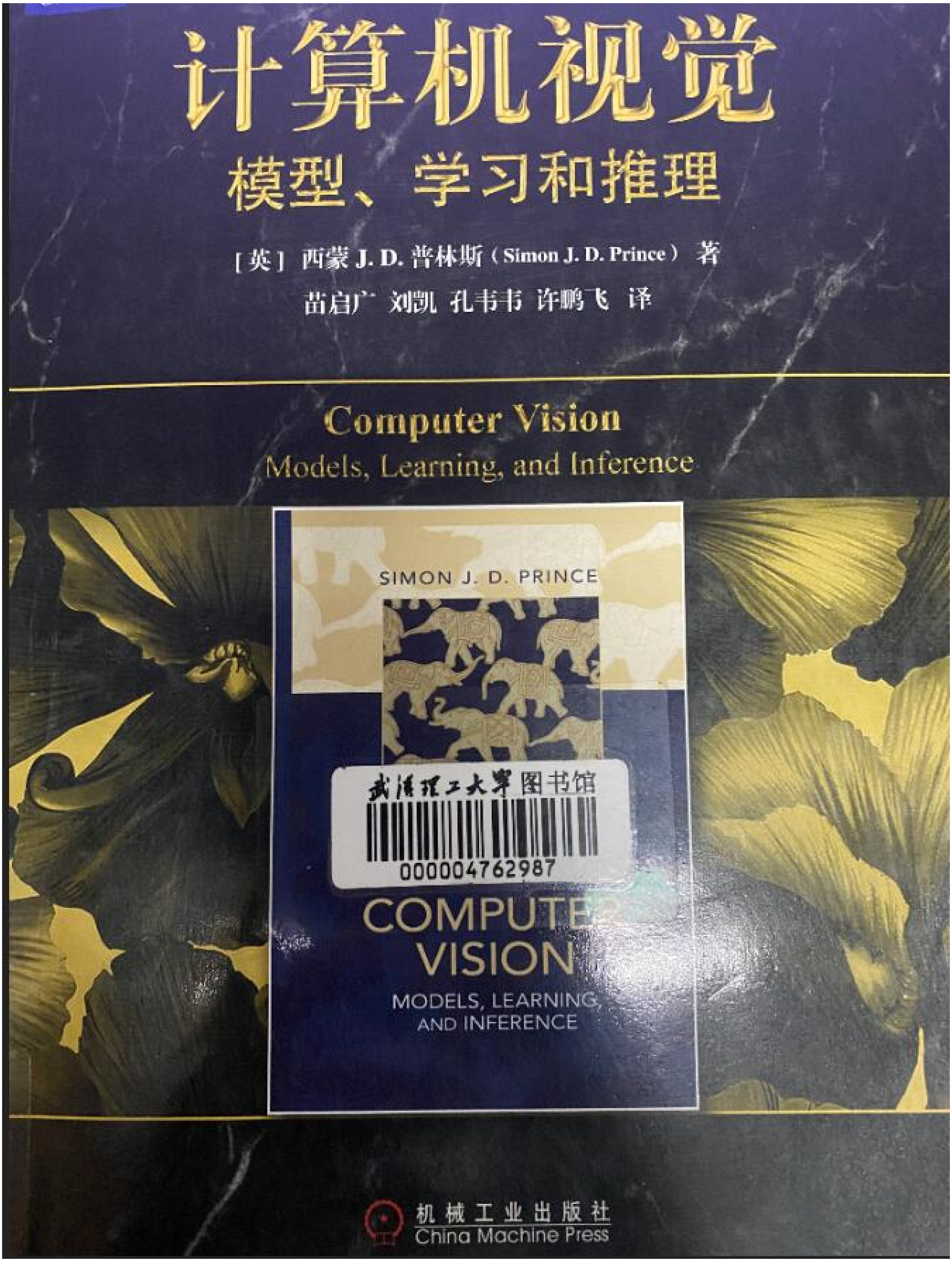
Table Of Contents：

今日工作

1、阅读书籍《计算机视觉 模型、学习和推理》的1~3章。

> 绪论
> 概率概述
> 常用概率分布



概率概述

书中提到，几乎所有的计算机视觉模型可以在概率范围内解释。概率是适合描述计算机视觉问题的语言。

目前，对这个论述还感到一些疑惑。

随机变量

随机变量可以是**离散**的或**连续**的。离散随机变量的**取值集合**可以是有序或无序的，有序的：骰子点数，无序的：观察天气的结果。可以是有限的或无线的，有限的：从52张排抽一张排的结果，无限的：下一班火车上的人数（理论上无限）。同样，连续随机变量的取值**取值范围**可以是有限或无限的。

对于离散随机变量，我们可以用描述每一个取值对应的可能性。

对于连续随机变量，由于一个区间内有无数个点，我们使用**概率密度函数(PDF)****描述变量的**概率分布**。某点的概率密度的值可以大于1，但概率密度函数的积分总是1，某点对应的概率密度值越大，说明其可能性越高。

随机变量的分布函数

定义：

设X是一个随机变量，x是任意实数，函数

$$F(x)=P\{X\leqslant x\}$$

称为X的**分布函数。**

即随机变量在(-∞, x]内取值的概率。

连续随机变量的概率密度

连续型随机变量的概率密度

定义：

对于**随机变量**X的**分布函数** `F(x)` ，如果存在非负函数 `f(x)` ，使对于任意实数x，有

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$$

则称*X*为**连续型随机变量**，函数 `f(x)` 称为*X*的**概率密度函数。**

根据

$$f(x)=\lim_{\Delta x\to 0^+}\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{\Delta x\to 0^+}\frac{P|x<X\leqslant x+\Delta x|}{\Delta x}$$

上式表明， `f(x)` **不表示**随机变量*X*取*x*的概率，而是反映*X*在点*x*的概率分布的密集程度。

由于变量取值是连续的，在区间上我们可以取无数个点，因此具体到一个数时，一般情况下概率是很小的，对于连续随机变量的概率，我们要描述的对象应该是一个区间，是在这个区间内取值的概率。

联合概率

研究对象是多个变量。

使用 `Pr(x,y)` 表示*x*和*y*的联合概率分布，其中，*x*和*y*不是标量，而是*D* \ 1*的向量。

边缘化

对于二元变量的概率分布函数 `Pr(x, y)` ，我们可以通过积分得到*x*或*y*的单变量概率分布函数 `Pr(x)` 和 `Pr(y)` 。

$$\mathrm{Pr}(x)=\int\mathrm{Pr}(x,y)\mathrm{d}y$$
$$\mathrm{Pr}(y)=\int\mathrm{Pr}(x,y)\mathrm{d}x$$

这样得到的分布 `Pr(x)` 和 `Pr(y)` 称为**边缘分布**，这个过程叫做**边缘化。**

对于离散变量，积分操作替换为累加操作。

例如一个四元变量的概率分布，可以通过以下计算得到 `Pr(x, y)`

$$\Pr(x, y) = \sum_{w} \int \Pr(w, x, y, z) \mathrm{d}z$$

边缘化可以理解为忽略被边缘的变量的影响考虑剩余变量的取值概率。

条件概率

$$\Pr(x \mid y) = \frac{\Pr(x, y)}{\Pr(y)}$$

$$\Pr(x, y) = \Pr(x \mid y) \Pr(y)$$

根据对称性，有

对于两个以上变量，有

贝叶斯公式

对于二元变量，贝叶斯公式根据上述条件概率中由对称称性得到的两个公式进行联立得到，公式为

其中 `Pr(y|x)` (等式左边)，叫做**后验概率**，代表给定*x*下*y*的概率；

`Pr(y)` ，叫做**先验概率**，表示不考虑*x*时，*y*的概率；

`Pr(x|y)` ，叫做**似然性**；

`Pr(x)` ，是**证据**。

期望

主要是一些特别函数的期望含义：

常用概率分布

对于一组变量*x*，其有对应的概率分布，概率分布的选择取决于建模数据*x*的定义域

每种分布都有对应的参数，这些参数决定概率分布的形状。

而分布的参数，又与一个分布联系。

共轭性

如上图，**贝塔分布**可以表征**伯努利分布**中**参数的概率**，与之相似，狄利克雷分布可表征分类分布参数的分布，同样的类比关系也适用于正态逆伽马分布与一元正态分布、正态逆维希特 分布与多元正态分布之间。

对于共轭性，目前的认识还不足。

在上述的配对关系中，前一个分布是后一个分布的共轭，例如**贝塔分布**与**伯努利分布**共轭。当把一个分布与其共轭分布相乘时，结果正比于一个新的分布，它与**共轭的形式**相同：

注意，共轭指的是参数的分布函数。

总结

今天由于课程较多，学习内容有限。

为了加强理论知识，对计算机视觉有全面的、深入的理解，今后将继续阅读这本书籍。

通过今天的学习，复习了概率的相关知识。

学习了**共轭性**这个新内容。但是目前对这个概念还不是很了解，书中提到

> 在学习(拟合分布)和评估模型(评估在拟合分布下新数据的概率)的过程中会用到分布的乘积，因此共轭关系很重要。共轭关系意味着这些乘积可以闭式求解。

对用到这个关系时的情景还没有清晰的认识。

> 使用概率分布可以描述全局状态和图像数据。为此已经给出了四个分布（伯努利分布， 分类分布、一元正态分布、多元正态分布）。还给出了另外四个分布（贝塔分布、狄利克雷分布、正态逆伽马分布、正态逆维希特分布)，可以用于描述上一组分布的参数的概率分布，因此它们可以描述拟合模型的不确定性。这 4 对分布有特殊关系：第二组中的每个分布是对应的第一组的共轭。正如我们看到的，共轭关系可以更容易地拟合观测数据并在拟合分布模型下评估新的数据。

那么，一个图像如何使用概率分布进行描述呢？有待通过后续的学习加深理解。