[复习]概率统计 / Probabilities and Statistics Refresher

翻译&校正 | 韩信子@ShowMeAI　　编辑 | 南乔@ShowMeAI　　原文作者 | https://stanford.edu/~shervine　　本节原文超链

[1]概率和组合简介 / Introduction to Probability and Combinatorics

样本空间 Sample space

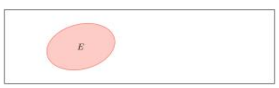
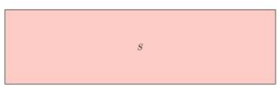
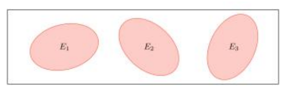
一个实验的所有可能结果的集合称为实验的样本空间，记作 S。

事件 Event

样本空间的任何子集 E 被称为一个事件。即，一个事件是一个包含可能结果的集合。如果该实验的结果包含在 E 内，那么称 E 发生。

概率论公理 Axioms of probability

对每个事件 E，记 $P(E)$ 为事件 E 出现的概率：

$0 \leq P(E) \leq 1$	$P(S) = 1$	$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$
		
概率在 0 到 1 之间（包含 0，1）	总体事件集合中至少 1 个事件出现了的概率是 1	互相独立的事件 $E_1, \ldots, E_n$满足上述公式

排列 Permutation

一个排列是从n个对象的池子中抽取r个对象进行排列（考虑顺序）。排列的数目为：

$$P(n,r) = \frac{n!}{(n-r)!}$$

组合 Combination

一个组合是从n个对象的池子中抽取r个对象（无序）。组合的数目为：

$$C(n,r) = \frac{P(n,r)}{r!} = \frac{n!}{r!\,(n-r)!}$$

备注：对于 $0 \leq r \leq n$， 有 $P(n,r) \geq C(n,r)$。

[2]条件概率 / Conditional Probability

贝叶斯法则 Bayes' rule

对事件 A 和 B 满足 $P(B) > 0$， 有：

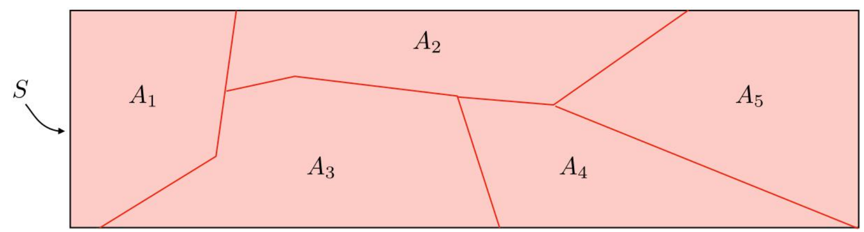
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

备注： $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$。

划分 Partition

令对所有 i， $A_i = \{A_i\}$。令 $\{A_i, i \in [\![1,n]\!]\}$ ，对所有 i， $A_i \neq \varnothing$ ，称 $\{A_i\}$ 为一个划分：

$$\forall i \neq j,\ A_i \cap A_j = \emptyset \quad 和 \quad \bigcup_{i=1}^{n} A_i = S$$



备注：对任意在样本空间中的事件 B， $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$ 。

贝叶斯法则的扩展形式 Extended form of Bayes' rule

令 $\{A_i, i \in [\![1,n]\!]\}$ 为样本空间的一个划分，则有：

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

独立 Independence

当且仅当两个事件 A 和 B 是独立的，有：$P(A \cap B) = P(A)P(B)$。

[3]随机变量 / Random Variables

3.1 Definitions

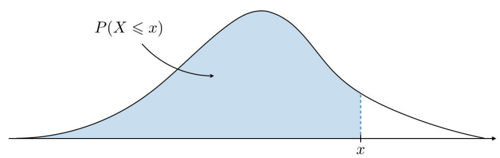
随机变量 Random variable

一个随机变量（记作 X）是一个函数，将一个样本空间中的每个元素映射到一个实值。

累积分布函数 CDF

累积分布函数（Cumulative distribution function，CDF）F 是单调不递减的，且 $\lim_{x\to-\infty} F(x) = 0$ 和 $\lim_{x\to+\infty} F(x) = 1$。F(x) 定义如下：

$$F(x) = P(X \le x)$$



备注：P(a < X ≤ B) = F(b) − F(a) 。

概率密度函数 PDF

概率密度函数（Probability density function，PDF）f，表示 X 在两个相邻随机变量的实现间取值的概率。

PDF 和 CDF 的关系 Relationships involving the PDF and CDF

下表总结了二者在离散和连续场景下的重要性质：

类型	CDF F	PDF f	PDF 的性质
离散	$F(x) = \sum_{x_i \le x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \le f(x_j) \le 1$ 和 $\sum_j f(x_j) = 1$
连续	$F(x) = \int_{-\infty}^{x} f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \ge 0$ 和 $\int_{-\infty}^{+\infty} f(x)dx = 1$

分布的期望和矩 Expectation and Moments of the Distribution

下表总结了期望值 $E[X]$、一般期望值 $E[g(X)]$、第 k 阶矩 $E[X^k]$ 和特征函数 $\psi(\omega)$ 在离散和连续场景下的表达式：

类型	期望值 E[X]	一般期望值 E[g(X)]	第 k 阶矩 $E[X^k]$	特征函数 $\psi(\omega)$
离散	$\sum_{i=1}^{n} x_i f(x_i)$	$\sum_{i=1}^{n} g(x_i) f(x_i)$	$\sum_{i=1}^{n} x_i^k f(x_i)$	$\sum_{i=1}^{n} f(x_i) e^{i\omega x_i}$
连续	$\int_{-\infty}^{+\infty} x f(x)dx$	$\int_{-\infty}^{+\infty} g(x)f(x)dx$	$\int_{-\infty}^{+\infty} x^k f(x)dx$	$\int_{-\infty}^{+\infty} f(x)e^{i\omega x}dx$

方差 Variance

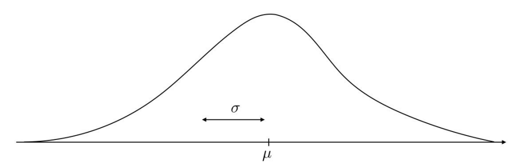
随机变量的方差通常记作$Var(X)$或σ^2，是分布函数的扩散性的一个度量函数。定义：

$$\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

标准差 Standard deviation

随机变量的标准差，通常记作σ，是分布函数扩散性的一个和实际随机变量值单位相当的度量函数。定义为：

$$\sigma = \sqrt{\mathrm{Var}(X)}$$ 。



随机变量的变换 Transformation of random variables

令变量 X 和 Y 由某个函数联系在一起。记 f_X 和 f_Y 分别为 X 和 Y 的分布函数：

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

莱布尼兹积分法则 Leibniz integral rule

令 g 为 x 和 c 的函数，a，b 是可能依赖于 c 的边界。

$$\frac{\partial}{\partial c}\left(\int_a^b g(x)dx\right) = \frac{\partial b}{\partial c} \cdot g(b) - \frac{\partial a}{\partial c} \cdot g(a) + \int_a^b \frac{\partial g}{\partial c}(x)dx$$

3.2 概率分布/ Probability Distributions

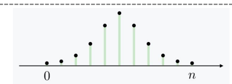
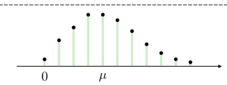
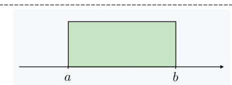
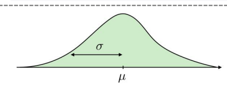
切比雪夫不等式 Chebyshev's inequality

随机变量 X 的期望值为 μ。对 $k,\sigma>0$，下列不等式成立：

$$P(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2}$$

主要分布 Main distributions

这里是主要需要记住的分布

	分布	概率密度 PDF	特征函数 $\psi(\omega)$	期望 E[X]	方差 Var(X)	图示
离散	$X\sim\mathcal{B}(n,p)$	$\binom{n}{x}p^xq^{n-x}$	$(pe^{i\omega}+q)^n$	np	npq	
	$X\sim\text{Po}(\mu)$	$\frac{\mu^x}{x!}e^{-\mu}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ	
连续	$X\sim\mathcal{U}(a,b)$	$\frac{1}{b-a}$	$\frac{e^{i\omega b}-e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
	$X\sim\mathcal{N}(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$e^{i\omega\mu-\frac{1}{2}\omega^2\sigma^2}$	μ	σ^2	
	$X\sim\text{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{1-\frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	

[4]联合分布随机变量 / Jointly Distributed Random Variables

边缘密度和累积分布 Marginal density and cumulative distribution

由联合密度概率函数 f_{XY}，可得：

	边缘密度函数 Marginal density function	累积函数 Cumulative function
离散	$f_X(x_i)=\sum_j f_{XY}(x_i,y_j)$	$F_{XY}(x,y)=\sum_{x_i\leq x}\sum_{y_j\leq y}f_{XY}(x_i,y_j)$
连续	$f_X(x)=\int_{-\infty}^{+\infty}f_{XY}(x,y)dy$	$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(x',y')dx'dy'$

条件密度 Conditional density

X 关于 Y的条件密度通常记作 $f_{X|Y}$， 定义：

$$f_{X|Y}(x)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

独立性 Independence

当两个随机变量 X 和 Y 满足如下特性时，称其为互相独立的：

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$

协方差 Covariance

两个随机变量 X 和 Y 的协方差，记作 σ_{XY}^2 或者更常见的 $\text{Cov}(X,Y)$， 定义如下：

$$\text{Cov}(X,Y)\triangleq\sigma_{XY}^2=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\mu_Y$$

相关性 Correlation

σ_X，σ_Y 为 X 和 Y 的标准差，ρ_{XY}为随机变量 X 和 Y 的相关性，其定义如下：

$$\rho_{XY}=\frac{\sigma_{XY}^2}{\sigma_X\sigma_Y}$$

备注： 对任何随机变量 X，Y， $\rho_{XY}\in[-1,1]$。 如果 X 和 Y 独立，$\rho_{XY}=0$ 。

[5]参数估计 / Parameter Estimation

5.1 Definitions

随机采样 Random sample

$X_1,...,X_n$ 是 n 个和 X 独立同分布的随机变量，随机采样是这些随机变量的集合。

预估器 Estimator

预估器是一个函数，用来推断一个统计模型中未知参数值。

偏差 Bias

估计器 $\hat{\theta}$ 的偏差，定义为 $\hat{\theta}$ 分布的期望值和真实值间的差距，即：

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

备注： 当 $E[\hat{\theta}] = \theta$，估计器被称为无偏的。

5.2 均值估计/ Estimating the Mean

样本均值 Sample mean

会用样本统计量（可以理解为随机抽样）来估计总体参数，比如总体均值 μ ，把样本均值记作 $\overline{X}$ ，可以通过如下公式计算得到：

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

备注： 样本均值是无偏的， i.e $E[\overline{X}] = \mu$。

中心极限定理 Central Limit Theorem

令随机采样 $X_1, \ldots, X_n$ 满足均值为 μ、方差为 σ^2的分布，则有：

$$\overline{X} \underset{n\to+\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

5.3 方差估计/ Estimating the variance

样本方差 Sample variance

会用样本统计量（可以理解为随机抽样）来估计总体参数[比如总体方差 σ^2]，把样本方差记作 s^2 或者 $\hat{\sigma}^2$，可以通过如下公式计算得到：

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

备注： 样本方差是无偏的， i.e $E[s^2] = \sigma^2$。

样本方差与卡方的关系 Chi-Squared relation with sample variance

令 s^2 表示随机样本的样本方差，有如下公式：

$$\frac{s^2(n-1)}{\sigma^2} \sim \chi_{n-1}^2$$

Awesome AI Courses Notes Cheat Sheets

Machine Learning CS229 Stanford	Deep Learning CS230 Stanford	Natural Language Processing CS224n Stanford	Computer Vision CS231n Stanford	Deep Reinforcement Learning CS285 UC Berkeley	Neural Networks for NLP CS11-747 CMU	DL for Self-Driving Cars 6.S094 MIT	... ...

是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 TOP20+门 AI 课程，旨在为读者和学习者提供一整套高品质中文速查表，可以点击【这里】查看。

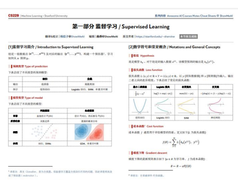
斯坦福大学（Stanford University）的 Machine Learning（CS229）和 Deep Learning（CS230）课程，是本系列的第一批产出。

本批两门课程的速查表由斯坦福大学计算机专业学生 Shervine Amidi 总结整理。原速查表为英文，可点击【这里】查看，ShowMeAI 对内容进行了翻译、校对与编辑排版，整理为当前的中文版本。

有任何建议和反馈，也欢迎通过下方渠道和我们联络 (*￣3￣)

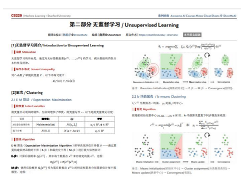
CS229 | Machine Learning @ Stanford University

监督学习 Supervised Learning



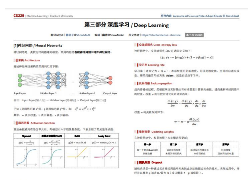
中文速查表链接

无监督学习 Unsupervised Learning



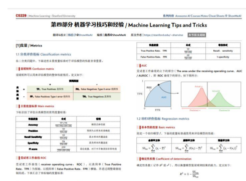
中文速查表链接

深度学习 Deep Learning



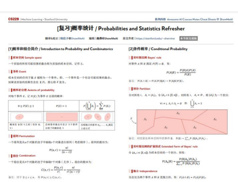
中文速查表链接

机器学习技巧和经验 Tips and Tricks



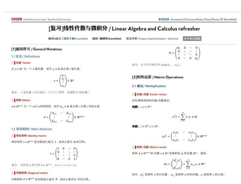
中文速查表链接

概率统计 Probabilities /Statistics



中文速查表链接

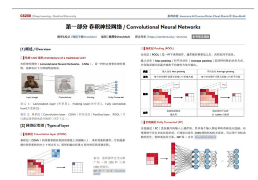
线性代数与微积分 Linear Algebra and Calculus



中文速查表链接

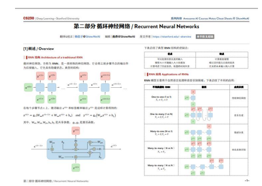
CS230 | Deep Learning @ Stanford University

卷积神经网络 CNN



中文速查表链接

循环神经网络 RNN



中文速查表链接

深度学习技巧与建议 Tips and Tricks

中文速查表链接

GitHub

ShowMeAI

https://github.com/ShowMeAI-Hub/

ShowMeAI 研究中心

扫码回复"速查表"

下载最新全套资料