在线阅读！！机器学习数学精华：概率论与数理统计

原创 机器学习初学者 机器学习初学者 2019-11-01 19:31

机器学习，需要一定的数学基础，需要掌握的数学基础知识特别多，如果从头到尾开始学，估计大部分人来不及，我建议先学习最基础的数学知识，基础知识可以分为**高等数学、线性代数、概率论与数理统计**三部分，我整理了相关数学基础资料：

源文件下载：

https://github.com/fengdu78/Data-Science-Notes/tree/master/0.math

内容简介

一、斯坦福大学CS229数学基础

这是斯坦福大学 CS 229 机器学习课程的基础材料，是斯坦福各大人工智能课程的数学基础，对人工智能课程做了优化，强烈推荐！！

我们对原始教程进行了翻译，翻译版本做成了在线阅读版本。

（点击查看：1.**线性代数**，2.**概率论**）

二、国内大学的数学基础教材精华

这个是我考研考博时候整理的中文教材的资料，分为高等数学、线性代数、概率论与数理统计三部分，我把和机器学习相关的数学知识进行了整理，进行公布。

本文是概率论和数理统计部分，建议收藏慢慢看。

概率论和数理统计

随机事件和概率

1.事件的关系与运算

(1) 子事件：$A \subset B$，若A发生，则B发生。

(2) 相等事件：$A = B$，即$A \subset B$，且$B \subset A$ 。

(3) 和事件：$A \bigcup B$（或$A + B$），A与B中至少有一个发生。

(4) 差事件：$A - B$，A发生但B不发生。

(5) 积事件：$A \bigcap B$（或AB），A与B同时发生。

(6) 互斥事件（互不相容）： $A\bigcap B$=$\varnothing$。

(7) 互逆事件（对立事件）： $A\bigcap B=\varnothing,A\bigcup B=\Omega,A=\bar{B},B=\bar{A}$

2.运算律(1) 交换律：$A\bigcup B=B\bigcup A,A\bigcap B=B\bigcap A$ (2) 结合律：$(A\bigcup B)\bigcup C=A\bigcup(B\bigcup C)$ (3) 分配律：$(A\bigcap B)\bigcap C=A\bigcap(B\bigcap C)$

3.德.摩根律

$\overline{A\bigcup B}$= $\bar{A}\bigcap\bar{B}$ $\overline{A\bigcap B}$= $\bar{A}\bigcup\bar{B}$

4.完全事件组

$A_1A_2\cdots A_n$两两互斥，且和事件为必然事件，即$A_i\bigcap A_j=\varnothing,i\neq j,\bigcup^n{}_{i=1}\ =\Omega$

5.概率的基本公式(1)条件概率: $P(B|A)=\frac{P(AB)}{P(A)}$,表示$A$发生的条件下，$B$发生的概率。

(2)全概率公式： $P(A)=\sum\limits_{i=1}^{n}P(A|B_i)P(B_i),B_iB_j=\varnothing,i\neq j,\bigcup^n{}_{i=1}\ B_i=\Omega$

(3) Bayes 公式：

$P(B_j|A)=\frac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{n}P(A|B_i)P(B_i)},j=1,2,\cdots,n$ 注：上述公式中事件B_i的个数可为可列个。

(4)乘法公式： $P(A_1A_2)=P(A_1)P(A_2|A_1)=P(A_2)P(A_1|A_2)$
$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$

6.事件的独立性

(1)A与B相互独立

$\Leftrightarrow P(AB)=P(A)P(B)$

(2)A，B，C两两独立 $\Leftrightarrow P(AB)=P(A)P(B)$;$P(BC)=P(B)P(C)$;$P(AC)=P(A)P(C)$;

(3)A，B，C相互独立 $\Leftrightarrow P(AB)=P(A)P(B)$; $P(BC)=P(B)P(C)$; $P(AC)=P(A)P(C)$; $P(ABC)=P(A)P(B)P(C)$

7.独立重复试验

将某试验独立重复n次，若每次实验中事件 A 发生的概率为p，则n次试验中A发生k次的概率为： $P(X=k)=C_n^kp^k(1-p)^{n-k}$

8.重要公式与结论 (1)$P(\bar{A})=1-P(A)$

(2)$P(A\bigcup B)=P(A)+P(B)-P(AB)$
$P(A\bigcup B\bigcup C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$

(3)$P(A-B)=P(A)-P(AB)$

(4)$P(A\bar{B})=P(A)-P(AB), P(A)=P(AB)+P(A\bar{B}),$
$P(A\bigcup B)=P(A)+P(\bar{A}B)=P(AB)+P(A\bar{B})+P(\bar{A}B)$

(5)条件概率$P(\centerdot|B)$满足概率的所有性质， 例如：. $P(\bar{A}_1|B)=1-P(A_1|B)$
$P(A_1\bigcup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$ $P(A_1A_2|B)=P(A_1|B)P(A_2|A_1B)$

(6)若$A_1,A_2,\cdots,A_n$相互独立，则$P(\bigcap\limits_{i=1}^{n}A_i)=\prod\limits_{i=1}^{n}P(A_i), P(\bigcup\limits_{i=1}^{n}A_i)=\prod\limits_{i=1}^{n}(1-P(A_i))$

(7)互斥、互逆与独立性之间的关系： A与B互逆$\Rightarrow$ A与B互斥，但反之不成立，A与B互斥（或互逆）且均非零概率事件$\Rightarrow$$A$与$B$不独立.

(8)若$A_1,A_2,\cdots,A_m,B_1,B_2,\cdots,B_n$相互独立，则$f(A_1,A_2,\cdots,A_m)$与$g(B_1,B_2,\cdots,B_n)$也相互独立，其中$f(\centerdot),g(\centerdot)$分别表示对相应事件做任意事件运算后所得的事件，另外，概率为 1（或 0）的事件与任何事件相互独立.

随机变量及其概率分布

1.随机变量及概率分布

取值带有随机性的变量，严格地说是定义在样本空间上，取值于实数的函数称为随机变量，概率分布通常指分布函数或分布律

2.分布函数的概念与性质

定义： $F(x)=P(X\le x),-\infty<x<+\infty$

性质：(1)$0\le F(x)\le 1$

(2) $F(x)$单调不减

(3) 右连续$F(x+0)=F(x)$

(4) $F(-\infty)=0,F(+\infty)=1$

3.离散型随机变量的概率分布

$P(X=x_i)=p_i,i=1,2,\cdots,n,\cdots \qquad p_i\ge 0,\sum_{i=1}^{\infty}p_i=1$

4.连续型随机变量的概率密度

概率密度$f(x)$;非负可积，且:

(1)$f(x) \geq 0,$

(2)$\int_{-\infty}^{+\infty} f(x)dx = 1$

(3)x为$f(x)$的连续点，则:

$f(x) = F'(x)$分布函数$F(x) = \int_{-\infty}^{x} f(t)dt$

5.常见分布

(1) 0-1 分布:$P(X=k) = p^k(1-p)^{1-k}, k = 0,1$

(2) 二项分布:$B(n,p)$： $P(X=k) = C_n^k p^k(1-p)^{n-k}, k = 0,1,\cdots,n$

(3) **Poisson**分布:$p(\lambda)$： $P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \lambda > 0, k = 0,1,2\cdots$

(4) 均匀分布$U(a,b)$： $f(x) = \{ \begin{matrix} \frac{1}{b-a}, a < x < b \\ 0, \end{matrix}$

(5) 正态分布:$N(\mu,\sigma^2) : \varphi(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \sigma > 0, \infty < x < +\infty$

(6)指数分布:$E(\lambda) : f(x) = \{ \begin{matrix} \lambda e^{-\lambda\ x}, x > 0, \lambda > 0 \\ 0, \end{matrix}$

(7)几何分布:$G(p) : P(X=k) =(1-p)^{k-1}p, 0 < p < 1, k = 1,2,\cdots.$

(8)超几何分布: $H(N,M,n) : P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0,1,\cdots, min(n,M)$

6.随机变量函数的概率分布

(1)离散型：$P(X=x_1) = p_i, Y = g(X)$

则: $P(Y=y_j) = \sum_{g(x_i)=y_i} P(X=x_i)$

(2)连续型：X~$f_X(x), Y = g(x)$

则:$F_y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{g(x)\leq y} f_x(x)dx,\quad f_Y(y) = F_Y'(y)$

7.重要公式与结论

(1) $X \sim N(0,1) \Rightarrow \varphi(0) = \frac{1}{\sqrt{2\pi}}, \Phi(0) = \frac{1}{2}, \Phi(-a) = P(X \leq -a) = 1 - \Phi(a)$

(2) $X \sim N\left(\mu,\sigma^2\right) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1), P(X \leq a) = \Phi(\frac{a-\mu}{\sigma})$

(3) $X \sim E(\lambda) \Rightarrow P(X > s+t|X > s) = P(X > t)$

(4) $X \sim G(p) \Rightarrow P(X = m+k|X > m) = P(X=k)$

(5) 离散型随机变量的分布函数为阶梯间断函数；连续型随机变量的分布函数为连续函数，但不一定为处处可导函数。

(6) 存在既非离散也非连续型随机变量。

多维随机变量及其分布

1.二维随机变量及其联合分布

由两个随机变量构成的随机向量(X, Y)， 联合分布为$F(x, y)=P(X \leq x, Y \leq y)$

2.二维离散型随机变量的分布

(1) 联合概率分布律 $P\{X=x_i, Y=y_j\}=p_{ij}; i, 1$

(2) 边缘分布律 $p_{i\cdot}=\sum_{j=1}^{\infty} p_{ij}, i=1,2, \cdots$ $p_{\cdot j}=\sum_{i}^{\infty} p_{ij}, j=1,2, \cdots$

(3) 条件分布律 $P\{X=x_i|Y$ $P\{Y=y_j|X$

3. 二维连续性随机变量的密度

(1) 联合概率密度$f(x, y)$：

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) d x d y=1$

(2) 分布函数：$F(x, y)=\int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) d u d v$

(3) 边缘概率密度： $f_X(x)=\int_{-\infty}^{+\infty} f(x, y) d y$ $f_Y(y)=\int_{-\infty}^{+\infty} f(x, y) d x$

(4) 条件概率密度：$f_{X|Y}(x|y)=\frac{f(x, y)}{f_Y(y)}$ $f_{Y|X}(y|x)=\frac{f(x, y)}{f_X(x)}$

4.常见二维随机变量的联合分布

(1) 二维均匀分布：$(x, y) \sim U(D)$,$f(x, y)=\begin{cases}\frac{1}{S(D)},(x, y) \in D \\ 0, \text{其他}\end{cases}$

(2) 二维正态分布：$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$,$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$f(x, y)=\frac{1}{2 \pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} . \exp \left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2 \rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1 \sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

5.随机变量的独立性和相关性

X和Y的相互独立:$\Leftrightarrow F(x, y)=F_X(x) F_Y(y)$:

$\Leftrightarrow p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$（离散型） $\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$（连续型）

X和Y的相关性：

相关系数$\rho_{XY}=0$时，称X和Y不相关， 否则称X和Y相关

6.两个随机变量简单函数的概率分布

离散型： $P(X=x_i,Y=y_i)=p_{ij},Z=g(X,Y)$ 则：

$P(Z=z_k)=P\{g(X,Y)=z_k\}=\sum_{g(x_i,y_i)=z_k}P(X=x_i,Y=y_j)$

连续型： $(X,Y)\sim f(x,y),Z=g(X,Y)$ 则：

$F_z(z)=P\{g(X,Y)\leq z\}=\iint_{g(x,y)\leq z}f(x,y)dxdy$， $f_z(z)=F_z'(z)$

7.重要公式与结论

(1) 边缘密度公式： $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy,f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$

(2) $P\{(X,Y)\in D\}=\iint_D f(x,y)dxdy$

(3) 若(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 则有：

1. $X\sim N\left(\mu_1,\sigma_1^2\right),Y\sim N(\mu_2,\sigma_2^2)$.
2. X与Y相互独立$\Leftrightarrow\rho=0$，即X与Y不相关。
3. $C_1X+C_2Y\sim N(C_1\mu_1+C_2\mu_2,C_1^2\sigma_1^2+C_2^2\sigma_2^2+2C_1C_2\sigma_1\sigma_2\rho)$
4. X关于$Y=y$的条件分布为： $N(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2))$
5. Y关于$X=x$的条件分布为： $N(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2))$

(4) 若X与Y独立，且分别服从$N(\mu_1,\sigma_1^2),N(\mu_1,\sigma_2^2)$, 则：$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,0)$,

C_1X+C_2Y~$N(C_1\mu_1+C_2\mu_2,C_1^2\sigma_1^2C_2^2\sigma_2^2)$.

(5) 若X与Y相互独立，$f(x)$和$g(x)$为连续函数， 则$f(X)$和$g(Y)$也相互独立。

随机变量的数字特征

1.数学期望

离散型：$P\{X=x_i\}=p_i,E(X)=\sum_i x_ip_i$;

连续型：$X \sim f(x), E(X)=\int_{-\infty}^{+\infty} x f(x) d x$

性质：

(1) $E(C)=C, E[E(X)]=E(X)$

(2) $E(C_1 X+C_2 Y)=C_1 E(X)+C_2 E(Y)$

(3) 若X和Y独立，则$E(XY)=E(X)E(Y)$

(4)$[E(XY)]^2 \leq E(X^2)E(Y^2)$

2.方差：$D(X)=E[X-E(X)]^2=E(X^2)-[E(X)]^2$

3.标准差：$\sqrt{D(X)}$,

4.离散型：$D(X)=\sum_i [x_i-E(X)]^2 p_i$

5.连续型：$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x) d x$

性质：

(1) $D(C)=0, D[E(X)]=0, D[D(X)]=0$

(2) X与Y相互独立，则$D(X \pm Y)=D(X)+D(Y)$

(3) $D(C_1 X+C_2)=C_1^2 D(X)$

(4) 一般有
$D(X \pm Y)=D(X)+D(Y) \pm 2Cov(X, Y)=D(X)+D(Y) \pm 2\rho\sqrt{D(X)}\sqrt{D(Y)}$

(5) $D(X)<E(X-C)^2, C \neq E(X)$

(6) $D(X)=0 \Leftrightarrow P\{X=C\}=1$

6.随机变量函数的数学期望

(1) 对于函数$Y=g(x)$

X为离散型：$P\{X=x_i\}=p_i, E(Y)=\sum_i g(x_i) p_i$;

X为连续型：$X \sim f(x), E(Y)=\int_{-\infty}^{+\infty} g(x) f(x) d x$

(2) $Z=g(X, Y)$;$(X, Y) \sim P\{X=x_i, Y=y_j\}=p_{ij}$; $E(Z)=\sum_i \sum_j g(x_i, y_j) p_{ij}$
$(X, Y) \sim f(x, y)$;$E(Z)=\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) d x d y$

7.协方差

$Cov(X,Y)=E[(X-E(X)(Y-E(Y))]$

8.相关系数

$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$,$k$阶原点矩 $E(X^k)$; k阶中心矩 $E\left\{[X-E(X)]^k\right\}$

性质：

(1) $Cov(X,Y)=Cov(Y,X)$

(2) $Cov(aX,bY)=abCov(Y,X)$

(3) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

(4) $|\rho(X,Y)|\leq 1$

(5) $\rho(X,Y)=1\Leftrightarrow P(Y=aX+b)=1$，其中$a>0$

$\rho(X,Y)=-1\Leftrightarrow P(Y=aX+b)=1$，其中$a<0$

9.重要公式与结论

(1) $D(X)=E(X^2)-E^2(X)$

(2) $Cov(X,Y)=E(XY)-$

(3) $|\rho(X,Y)|\leq 1$,且 $\rho(X,Y)=1\Leftrightarrow P(Y=aX+b)=1$，其中$a>0$

$\rho(X,Y)=-1\Leftrightarrow P(Y=a$，其中$a<0$

(4) 下面 5 个条件互为充要条件：

$\rho(X,Y)=0\Leftrightarrow Cov(X,Y)=0\Leftrightarrow E(X,Y)=E(X)E(Y)\Leftrightarrow D(X+Y)=D(X)+D(Y)$
$\Leftrightarrow D(X-Y)=D(X)+D(Y)$

注：X与Y独立为上述 5 个条件中任何一个成立的充分条件，但非必要条件。

数理统计的基本概念

1.基本概念

总体：研究对象的全体，它是一个随机变量，用X表示。

个体：组成总体的每个基本元素。

简单随机样本：来自总体X的n个相互独立且与总体同分布的随机变量$X_1, X_2 \cdots, X_n$，称为容量为n的简单随机样本，简称样本。

统计量：设$X_1, X_2 \cdots, X_n$,是来自总体X的一个样本，$g(X_1, X_2 \cdots, X_n)$）是样本的连续函数，且$g()$中不含任何未知参数，则称$g(X_1, X_2 \cdots, X_n)$为统计量。

样本均值：$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

样本方差：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

样本矩：样本k阶原点矩：$A_k = \frac{1}{n}\sum^{n} X^k$

样本k阶中心矩：$B_k = \frac{1}{n}\sum^{n}(X_i - \overline{X})^k$

2.分布

χ^2分布：$\chi^2 = X_1^2 + X_2^2 + \cdots +$，其中$X_1, X_2 \cdots, X_n$,相互独立，且同服从$N(0,1)$

t分布：$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$，其中$X \sim N(0,1), Y \sim \chi^2(n)$,且$X$，$Y$ 相互独立。

F分布：$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$，其中$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$,且$X$，$Y$相互独立。

分位数：若$P(X \leq x_\alpha) = \alpha$,则称$x_\alpha$为$X$的$\alpha$分位数

3.正态总体的常用样本分布

(1) 设$X_1, X_2 \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的样本，

$\overline{X} \quad 1 \sum^{n} \quad X \quad \sigma 2$ 则：

1. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 或者$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
2. $(n-1)S^2 \quad 1 \sum^{n} ($
3. $\frac{1}{\ } \sum^{n} (X_i$

4) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

4.重要公式与结论

(1) 对于$\chi^2 \sim \chi^2(n)$，有 $E(\chi^2(n)) = n$

(2) 对于$T \sim t(n)$，有$E(T) = 0, D(T) = \frac{n}{n-2}(n > 2)$；

(3) 对于$F \sim F(m,n)$，有 $\frac{1}{F} \sim F(n, m)$, $F_{\alpha/2}($

(4) 对于任意总体X，有 $E(\overline{X}) = E(X)$, $E(S^2) =$

END

关于本站

“机器学习初学者”公众号由是黄海广博士创建，黄博个人知乎粉丝23000+，github排名全球前110名（32000+）。本公众号致力于人工智能方向的科普性文章，为初学者提供学习路线和基础资料。原创作品有：吴恩达机器学习个人笔记、吴恩达深度学习笔记等。

往期精彩回顾

- 那些年做的学术公益-你不是一个人在战斗
- 适合初学者入门人工智能的路线及资料下载
- 吴恩达机器学习课程笔记及资源（github标星12000+，提供百度云镜像）
- 吴恩达深度学习笔记及视频等资源（github标星8500+，提供百度云镜像）
- 《统计学习方法》的python代码实现（github标星7200+）

备注：加入本站微信群或者qq群，请回复“**加群**”

加入知识星球（4300+用户，ID：92416895），请回复“**知识星球**”

收录于合集 #机器学习的数学基础（在线阅读） 5

上一篇 · 在线阅读！！机器学习数学精华：线性代数