Mathematical basis

-PROMISE-

Tuesday 5th April, 2022

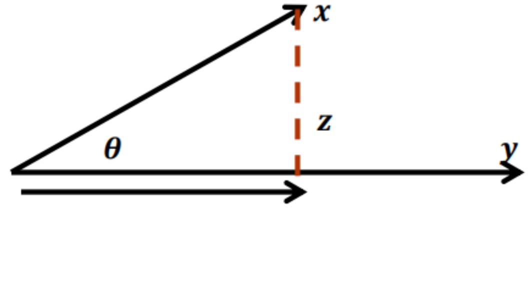
1 线性代数

1.1 向量 vector

- 对于向量 $x = (x_1\ x_2, ..., x_d)^d \in \mathbb{R}^d$
- 内积：$x^T y = y^T x = \sum_i^d x_i y_i$
- 向量长度 $\| x \| = \sqrt{x^T x}, \| x \|^2 = x^T x$
- $\| x \| = 1$ 的时候，x 是单位向量
- 正交：$x^T y = 0$ 时，x，y 互相垂直

1.2 内积，角度，投影

- $\| x \|$ 为向量长度，$\frac{x}{\|x\|}$ 为向量方向
- 向量夹角：$x^T y = \| x \| \cdot \| y \| \cdot \cos\theta, \| x^T y \| \leq \| x \| \cdot \| y \|$
- x 在 y 上的投影（projection）：
  方向：$\frac{y}{\|y\|}$, 长度：$\| x \| \cos\theta = \| x \| \frac{x^T y}{\|x\| \cdot \|y\|} = \frac{x^T y}{\|y\|}$
- 投影：
  $proj_y x = \frac{x^T y}{\|y\|^2} y$, $proj_y x \perp z$, $proj_y x + z = x$



1.3 柯西-施瓦茨不等式

1.3.1 Cauchy’s inequality

- $(\sum_{k=1}^{n} a_k b_k)^2 \leq (\sum_{k=1}^{n} a_k^2)(\sum_{k=1}^{n} b_k^2)$
  等号当且仅当存在固定实数 c，使得 $\forall k, a_k = cb_k$

1.3.2 Schwarz’s Inequality

- $[\int_a^b f(x)g(x)dx]^2 \leq [\int [f(x)]^2 dx][\int [g(x)]^2 dx]$
  等号当且仅当存在固定实数 c，使得 $\forall x \in [a,b], f(x) = cg(x)$

1.4 矩阵

- 矩阵 $m \times n$ 的矩阵，m 行 n 列
  m=n 时为方阵
  对角阵是只有对角线非零的方阵，其中对角线全是 1 就是单位阵

I_n T

- 方阵行列式的值
  $|X| = |X^T|$
  $|XY| = |X||Y|$
  $|\lambda X| = \lambda^n |X| (X : n \times n)$
- 逆矩阵
  $|X| \neq 0$ 则 X 可逆（|X| 写作 det(X)）
  $XX^{-1} = X^{-1}X = I_n$
  $(X^{-1})^{-1} = X, (\lambda X)^{-1} = \frac{1}{\lambda} X^{-1}$
  $(XY)^{-1} = Y^{-1}X^{-1}, (X^{-1})^T = (X^T)^{-1}$

1.5 方阵特征值

- $Ax = \lambda x$ λ x
  特征值和对角线的关系：$\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$, $det(A) = \prod_{i=1}^{n} \lambda_i$
  方阵的迹（trace）：$tr(A) = \sum_{i=1}^{n} a_{ii} = ??, tr(AB) = tr(BA)$

1.6 实对称矩阵

- 矩阵中每一项都是实数

特征值都是实数，特征向量都是实向量

特征值为 $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$

对应的特征向量记为 $\xi_1, \xi_2, ...,]xi_n$

特征向量互相垂直：$\xi_i^T \xi_j = 0 (i \neq j)$

$E = [\xi_1 \xi_2 ... \xi_n] \quad rank(E) = n$

1.7 实对称矩阵

- $X = \sum_{i=1}^{n} \lambda_i \xi_i \xi_i^T$
  属于矩阵的谱分布，约定 $\parallel \xi_i \parallel = 1$，则 E 实则正交矩阵
  $X = E \Lambda E^T$，$\Lambda \qquad \Lambda_{ii} = \lambda_i$
  $EE^T = E^T E = I$

1.8 正定，半正定

- 对称矩阵 A 是正定的当且仅当：
  $\forall x \neq 0, x^T A x = \sum_{i,j} x_i x_j A_{ij} > 0$, 此时特征值全是正数
  $\forall x \neq 0, x^T A x = \sum_{i,j} x_i x_j A_{ij} \geq 0$ 则 A 为半正定
  记为 $A \succ 0$ 或 $A \succcurlyeq 0$
  常用到二次型：$x^T A x$

1.9 矩阵求导

- 导数存在的时候：
  $\frac{\partial \mathbf{a}}{\partial x}$ 是一个向量，$(\frac{\partial \big(a)}{\partial x})_i = \frac{\partial a_i}{\partial x}$
  对于矩阵 $(\frac{\partial A}{\partial x})_{ij} = \frac{\partial A_{ij}}{\partial x}$
  $(\frac{\partial x}{\partial \mathbf{a}})_i = \frac{\partial x}{\partial a_i}$，$(\frac{\partial x}{\partial A})_{ij} = \frac{\partial a_i}{\partial A_{ij}}$，$(\frac{\partial \mathbf{a}}{\partial \mathbf{x}})_{ij} = \frac{\partial a_i}{\partial x_j}$
  用 Matrix Cookbook 查表

2 概率论

2.1 概率密度函数，分布函数

- PDF 和 CDF 的关系：p(x)=F'(x)

2.2 联合、条件分布、变换

- 联合：$P(X=x)$，$p(x)\geq 0\int p(x)dx=1$
- 条件：$p(x,y)=p(y)p(x|y)$，$p(x)=\int_y p(x,y)dy$(边界分布)
- 假设 $x=g(y)$，$p_Y(y)=p_X(x)|\frac{dx}{dy}|=p_X(g(y))|g'(y)|$

2.3 期望和方差

- 常规的省略，条件期望：$E(f(x)|Y=y)=\sum_x f(x)\cdot p(x|y)$

2.4 估计均值和协方差矩阵

- 训练样本：$x_1,x_2,...,x_n$
  均值估计 $x=\frac{1}{n}\sum_{i=1}^n x_i$
  Covariance 估计：$Cov(x)=\frac{1}{n}\sum_{i=1}^n(x_i-x)(x_i-x)^T$
  无偏估计：$Cov(x)=\frac{1}{n-1}\sum_{i=1}^n(x_i-x)(x_i-x)^T$

2.5 两随机变量的独立、相关

- (,) = (x) ( ) 则 X,Y 互相独立
  cov(X,Y)=E[(X-EX)(Y-EY)]=EX+EY-EXY
  $\rho_{XY}=\frac{cov(X,Y)}{\sqrt{(Var(X)Var(Y))}}$ $\rho_{XY}=0$ 时为不相关，$\rho_{XY}\pm 1$,
- 独立保证一定不相关，但不相关不一定独立

2.6 正态分布（高斯分布）

- 一维：$p(x)=(2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}exp-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}$
  多维:：$p(x)=(2\pi)^{-\frac{D}{2}}|\Sigma|^{-\frac{1}{2}exp-\frac{1}{2}(x-\mu)|\Sigma|^{-1}(x-\mu)}$, 其中 D： ,Σ μ：

，记为 $x\ N(\mu,\Sigma)$

- 正态分布中不相关等价于独立
  两个正态分布的加权和：aX+bY 也是正态分布
  条件分布：$x_a|x_b$ 也是高斯分布
  边际分布：$p(x_a)=\int p(x_a,x_b)dx_n$ 也是正态分布