Midterm review

Lec 1

马氏距离 $d(x,y)=\sqrt{(x-y)^T S^{-1}(x-y)}$.

$$E[X^TAX]=tr(A\Sigma)+\mu^TA\mu$$
$$E(||X-Y||_2^2)=2tr(\Sigma)$$
$$E(||X-Y||_\Sigma^2)=2trI=2p$$

Lec 2 线性代数

Projection x to y: $\frac{x'y}{y'y}y$.

矩阵乘向量的几何直观: 旋转+缩放+旋转 考虑谱分解，SVD分解

谱分解:

$$A=\sum_{i=1}^k\lambda_ie_ie_i^T=P\Lambda P'$$
$$A^{-1}=\sum_{i=1}^k\frac{1}{\lambda_i}e_ie_i^T=P\Lambda^{-1}P'$$
$$A^{-1/2}=\sum_{i=1}^k\frac{1}{\sqrt{\lambda_i}}e_ie_i^T=P\Lambda^{-1/2}P'$$

正定 = 特征值均大于0

马氏距离几何直观

$$Let\ x=\tilde{P}\tilde{\Lambda}y,\ s.t.\ var(y)=I$$
$$y'y=x'(\tilde{P}\tilde{\Lambda}^2\tilde{P}')^{-1}x$$
$$\Sigma=var(x)=\tilde{P}\tilde{\Lambda}^2\tilde{P}'$$
$$y'y=x'S^{-1}x$$

统计量几何表示

$$\bar{x}=\frac{1}{n}X'1_n$$
$$nS_n=X'(I-\frac{1}{n}1_n1_n')X$$
$$D=diag(S_n),R=D^{-1/2}S_nD^{-1/2}$$

广义样本方差: $\det S$

SVD分解:

$$A=U\Lambda V'$$
$$AA'=U\Lambda^2U',A'A=V\Lambda^2V'$$

不等式:

$b, d \in R^p$,$B \in R^{p\times p}$ is positive definite matrix. Then

$$b^T d \le (b^T Bb)(d^T Bd)$$

with equality if and only if $b = cB^{-1}d$ for some constant c.

$$\max_{x\neq 0} \frac{(x^T d)^2}{x^T Bx} = d^T B^{-1} d \text{ when } x = cB^{-1}d$$

Let B be a positive definite matrix with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$, associated with eigenvectors $e_1, e_2, \cdots, e_n$.

$$\max_{x\neq 0} \frac{x^T Bx}{x^T x} = \lambda_1, \text{ when } x = e_1$$
$$\max_{x\perp e_1,\cdots,e_k} \frac{x^T Bx}{x^T x} = \lambda_{k+1}, \text{ when } x = e_{k+1}$$

矩阵微积分

$$\frac{\partial}{\partial x} x^T Ax = (A + A^T)x$$
$$\frac{\partial}{\partial A}|A| = |A|A^{-1}$$
$$\frac{\partial}{\partial A} tr(AB) = B^T$$
$$\frac{\partial}{\partial A} tr(A^{-1}B) = -A^{-1}B^T A^{-1}$$

Lec 3 多元正态分布

$X \sim N_p(\mu, \Sigma), f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp[-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}]$

Contours of constant density:

$$(x-\mu)'\Sigma^{-1}(x-\mu) = c^2$$

These ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}e_i$, where e_i is an eigenvalue, i.e. $\Sigma e_i = \lambda_i e_i$.

多元正态的任意一组子变量仍符号多元正态，反过来不一定

$AX + d \sim N_p(A\mu + d, A\Sigma A')$.

若$Z \sim N_p(0, I)$, 则$\Sigma^{1/2}Z + \mu \sim N_p(\mu, \Sigma)$.

对于多元正态分布，独立=不相关

$X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

多元正态分布X, $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_p$.

$$\begin{aligned}(X-\mu)'\Sigma^{-1}(X-\mu) &= [(X-\mu)'\Sigma^{-1/2}][(X-\mu)\Sigma^{-1/2}]' \\ &= ZZ',\ where\ Z \sim N(0, I)\end{aligned}$$

多元正态分布的线性组合:

assume X_j mutually independent, $X_j \sim N_p(\mu_j, \Sigma), c_j \in R$.

$\sum_{i=1}^n c_i X_i \sim N(\sum_{i=1}^n c_i\mu_i, \sum_{i=1}^n c_i^2\Sigma)$

MLE: $\hat{\mu}=\bar{X}, \hat{\Sigma}=S_n=\frac{n-1}{n}S$

MLE的函数还是MLE

充分统计量$\bar{X}, S$, P28

$\bar{X}\sim N_p(\mu\frac{1}{n}\Sigma), (n-1)S\sim W_p(n-1,\Sigma)$,$\bar{X}, S$相互独立

Wishart distribution: $W_p(m,\Sigma)=\sum_{i=1}^m Z_jZ_j'$, where $Z_j\sim N_p(0,\Sigma)$.

properties: 可加性，P31

对于一组i.i.d的$\{X_j\}$

大数定律 $\bar{X}\to E(X_i)$

中心极限定理$\sqrt{n}(\bar{X}-\mu)$近似于$N_p(0,\Sigma)$, $n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)$近似于$\chi_p^2$, $n\gg p$.

Lec 4 单总体正态推断

对一组i.i.d.的$\{X_j\}$, $X_j\sim N_p(\mu,\Sigma)$.

Under H_0: $T^2=n(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0)\sim\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)$.

reject if $T^2>T^2(\alpha)=\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)$.

Hotelling's T^2 在线性变换下不变

$100(1-\alpha)\%$置信区域 $n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)\le T^2(\alpha)$

非正态假设下的大样本推断

$n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)>\chi_p^2(\alpha)$.

Simultaneous: $a'\bar{x}\pm\sqrt{\chi_p^2(\alpha)a'Sa/n}$

Simultaneous comparisons of component means: $(100(1-\alpha)\%)$

$\bar{x}_i\pm c\sqrt{\frac{s_{ii}}{n}}$

T^2: $c=\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}$

庞弗洛尼: $c=t_{n-1}(\frac{\alpha}{2m})$.

T^2区间比

Lec 5 双总体正态推断

1.成对比较

2.等方差正态假设

$T^2=(1/n_1+1/n_2)^{-1}(\bar{X}_1-\bar{X}_2-\delta_0)'S_{pooled}^{-1}(\bar{X}_1-\bar{X}_2-\delta_0)$

$S_{pooled}=\dfrac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}\sim\dfrac{W_p(n_1+n_2-2,\Sigma)}{n_1+n_2-2}.$

under $H_0, T^2\sim\dfrac{p(n_1+n_2-2)}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}$

Simultaneous: $a'(\bar{X}_1-\bar{X}_2)\pm\sqrt{a'(1/n_1+1/n_2)S_{pooled}a}$

对$\mu_1-\mu_2$的检验，拒绝假设起主要作用的线性组合: $S_p^{-1}(\bar{x}_1-\bar{x}_2)$

3.不等方差大样本

$(\bar{X}_1-\bar{X}_2-\delta_0)'(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2)^{-1}(\bar{X}_1-\bar{X}_2-\delta_0)\le\chi_p^2(\alpha).$

Simultaneous: $a'(\bar{x}_1-\bar{x}_2)\pm\sqrt{\chi_p^2(\alpha)a'(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2)a}.$

对$\mu_1-\mu_2$的检验，拒绝假设起主要作用的线性组合: $(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2)^{-1}(\bar{x}_1-\bar{x}_2)$

等方差检验

P39 textbook P237

独立性检验

正态性检验

Q-Qplot

Lec 6 PCA

总体主成分

Σ的特征值-特征向量对$(\lambda_1,e_1),(\lambda_2,e_2),\cdots,(\lambda_p,e_p)$, $\lambda_1\ge\cdots\ge\lambda_p\ge0$, 第j主成分$Y_j=e_j'X$.

$$var(Y_j)=\lambda_j, Cov(Y_i,Y_j)=0(i\ne j)$$
$$\sum_{i_1}^{p}Var(X_i)=\sum_{i=1}^{p}Var(Y_i)$$

$\rho_{Y_i,X_j}=\dfrac{e_{ij}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$

标准化后，主成分改变。

样本主成分

对样本协方差矩阵做特征分解即可。

主成分得分 $\hat{y}_i=\hat{e}_i'x$ 或 $\hat{y}_i=\hat{e}_i'(x-\bar{x})$. 可以看作样本$x$在$\hat{e}_i$方向上的投影，是一个实数