191220133 徐思源 本科生 计算机系 计算机[illegible]

6.1.(a) 设X是$m\times n$的矩阵，列秩为r，因此X的列空间维度为r，令$c_1, c_2 \cdots c_r$是X列空间的基，构成$m\times r$的矩阵$C=(c_1, c_2 \cdots c_r)$.

∵ A的列向量是C列向量线性组合构成的，

∴ 存在一个矩阵$R(r\times n)$ 使 $A=CR$.

∴ A的行向量是R行向量的线性组合

∴ A的行秩 $\le$ R的行秩 $\le r =$ A的列秩. ①

A的列秩 $= A^T$行秩 $\le A^T$列秩 $=$ A的行秩 ②

由①②知 A的行秩等于列秩 ∴ $\mathrm{rank}(X)=\mathrm{rank}(X^T)$.

(b) ∵ $\mathrm{rank}(A^T)=\mathrm{rank}(A)$ 且 $\mathrm{rank}(A^T)\le n$ $\mathrm{rank}(A)\le m$.

∴ $\mathrm{rank}(A)\le \min(m,n)$

(c) 类似(a)中 X的基为$a_1 \cdots a_s$，Y的基为$b_1 \cdots b_t$

∴ X行向量由$a_{1\sim s}$线性组合，Y行向量由$b_{1\sim t}$线性组成.

易知 X+Y 由$a_{1\sim s}$ $b_{1\sim t}$线性组合. ∴ $\mathrm{rank}(X+Y)\le s+t=\mathrm{rank}(X)+\mathrm{rank}(Y)$.

(d) 类似(c)中，XY中的行向量是 $a_1Y \sim a_sY$ 或 $Xb_1 \sim Xb_t$（两种表示）.

∴ $\mathrm{rank}(XY)\le \min(\mathrm{rank}(X), \mathrm{rank}(Y))$.

(e) ∵ $\mathrm{rank}(X)=\mathrm{rank}(X^T)$. 不失一般性，令$m\le n$.

∴ $\mathrm{rank}(X)=\mathrm{rank}(X^T)=m$. X行满秩，$X^T$列满秩.

∴ X, X^T可逆 ∴ $X=XX^T(X^T)^{-1}$ $\le \mathrm{rank}(XX^T)$.

同时 $\mathrm{rank}(XX^T)\le m=\mathrm{rank}(X)$.

∴ $\mathrm{rank}(X)=\mathrm{rank}(XX^T)$ 同理可得 $\mathrm{rank}(X)=\mathrm{rank}(XX^T)=\mathrm{rank}(X^TX)$.

(f) $\mathrm{rank}(XX^T)\le \min(\mathrm{rank}(X), \mathrm{rank}(X^T))=1$. XX^T秩为1.

∵ 存在可逆矩阵 Y 使 $Y^{-1}(XX^T)Y=\Lambda$（对角矩阵秩为非零值数量）.

∴ $\mathrm{rank}(X)=\mathrm{rank}(\Lambda)=$ 特征值数量.

6.4(a) 直接计算可知正确

(b) 保留v前k位，之后变为0.

(c) A变成一个上三角矩阵.

(d) 易知 $\det(M_k)=\det(M'_k(1,1))+0\cdot\det(M'_k(1,2))\cdots=1$（下三角）

$L=M_{n-1}M_{n-2}\cdots M_1=\begin{pmatrix} -\frac{v_2}{v_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ -\frac{v_n}{v_1} & -\frac{v_n}{v_{n-1}} & 1 \end{pmatrix}$ L为下三角 $\det L=1\times1\cdots\times1=1$

$(\det(M_k)=1)$.

(e) 由(d)知存在 $M_1\cdots M_{n-1}$ 使 $M_{n-1}\cdots M_1A=U$，其中 U 为上三角矩阵.

$M_{n-1}\cdots M_1$ 可逆 $\therefore A=LU$ $\left(L=(M_{n-1}\cdots M_1)^{-1}\right)$.

若A非奇异，设有另一组 $\lambda U'$ 使 $A=XY=LU$

$\Rightarrow \underset{\text{上三角}}{L^{-1}X}=\underset{\text{下三角}}{UY^{-1}}$ $\Rightarrow$ 必是单位矩阵

$\therefore L^{-1}=X$ $U=Y^{-1}$ 矛盾 $\therefore$ 是唯一的.

(f) 由(e)知 A有唯一的LU分解，易知L对角线为1.

令 $U=DX$ $\quad D=\begin{pmatrix} u_{11} & & \\ & \ddots & \\ & & u_{nn} \end{pmatrix}$ $\quad X=\begin{pmatrix} 1 & \frac{u_{12}}{u_{11}} & \cdots & \frac{u_{1n}}{u_{11}} \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$.

$A^T=A$ $\quad \therefore A^T=(LDX)^T=X^TD^TL^T=X^TDL^T=LDX=A$.

由(e)唯一性知 $X=L^T$ $\quad \therefore A=LDL^T$.

(g) 由f知 $A=LDL^T$. 设 $D=\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$ 则 $\sqrt{D}=\begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix}$. (易知 $\sqrt{D}\sqrt{D}=D$).

$\therefore A=LDL^T=L\sqrt{D}(L\sqrt{D})^T$

令 $G=L\sqrt{D}$ 则 $A=GG^T$

$\therefore$ A的 cholesky 分解也唯一.

T1 在github下载源码并make编译：

```
miracle@ubuntu:~/Desktop/libsvm-master$ make
g++ -Wall -Wconversion -O3 -fPIC -c svm.cpp
g++ -Wall -Wconversion -O3 -fPIC svm-train.c svm.o -o svm-train -lm
g++ -Wall -Wconversion -O3 -fPIC svm-predict.c svm.o -o svm-predict -lm
g++ -Wall -Wconversion -O3 -fPIC svm-scale.c -o svm-scale
miracle@ubuntu:~/Desktop/libsvm-master$
```

(b) i) Accurancy = 66.925% (2677/4000)

ii) Accurancy = 96.925% (3877/4000)

iii) Accurancy = 95.675% (3827/4000)

iv) Accurancy = 70.475% (2819/4000).

v) Accurancy = 96.875% (3875/4000)

(c) 有不平衡数据集 svmguide3.

默认参数下 Accurancy = 2.43902%.

提高频率较低类的权重，如调整为 2.5 和 0.6，则 Accurancy 上升为 75.6098%.

7.2 (a) $K_{HI}(x,y)=\min(x,y)$ 映射为 ϕ ∴ $K_{HI}(x,y)=\sum_{i=1}^{d}\min(x_i,y_i)=\phi_1+\cdots+\phi_d$.

∴存在一个映射 $f(x,y)=\phi_1+\cdots+\phi_n$ ∴合法.

(b) ∵X可以通过行列变换和Y互相转换 ∴Y是(半)正定则X(半)正定.

(c) 设 $P=\begin{pmatrix}1 & -1 & & 0\\ & \ddots & \ddots & \\ & & \ddots & -1\\ 0 & & & 1\end{pmatrix}$ ∴ $P^TXP=\begin{pmatrix}x_1 & & & 0\\ & x_2-x_1 & & \\ & & \ddots & \\ 0 & & & x_n-x_{n-1}\end{pmatrix}=\Lambda$

P^TXP是半正定的 ∴X是半正定的(∵P满秩 ∴ $X=(P^T)^{-1}\Lambda P^{-1}\rightarrow$ LDL的分解).

(d) 设 $x_i\le y_i$ 易知 $x_i\le\frac{2x_iy_i}{y_i+y_i}\le\frac{2x_iy_i}{x_i+y_i}$

∴ $K_{HI}(x,y)=\sum_{i=1}^{d}\min(x_i,y_i)\le\sum_{i=1}^{d}\frac{2x_iy_i}{x_i+y_i}=k_{\chi^2}(x,y)$.

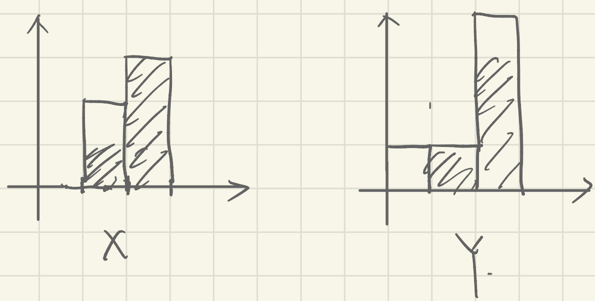
(e) 令 $f(x)=\sqrt{x}$

∴ $K_{HE}=\sum_{i=1}^{d}f(x_i)f(y_i)=f(x)^Tf(y)$. ∴是合法的, 映射为 f.

易知 $\frac{2x_iy_i}{x_i+y_i}\le\frac{2x_iy_i}{2\sqrt{x_iy_i}}=\sqrt{x_iy_i}$.

∴ $k_{\chi^2}(x,y)\le K_{HE}(x,y)$.

(f) 假设两数据分布图可以用直方图表示



可以理解为取两直方图的交集(数据上的).

↑

度量不一定相同

9.2 (a) 若 S_i 是 i 最近邻 k 的样本的集合, 则 $e_i=\|x_i-\sum_{s\in S_i}w_{is}x_s\|^2$ 且 $\sum_{s\in S_i}w_{is}=1$

$\sum_{i=1}^{n}e_i=\sum_{i=1}^{n}\|\sum_{s\in S_i}w_{is}x_i-\sum_{s\in S_i}w_{is}x_s\|^2$

$=\sum_{i=1}^{n}w_i^T\left(\sum_{s\in S_i}(x_i-x_s)(x_i-x_s)^T\right)w_i$

令 $Y_i=\sum_{s\in S_i}(x_i-x_j)(x_i-x_j)^T$ 则 $\sum_{i=1}^{n}e_i=\sum_{i=1}^{n}w_i^TY_iw_i$

由 $\sum_{s\in S_i}w_{ij}=w_i^Te_1=1$. ($e_1$为全1向量)

拉格朗日函数 $L(w,\lambda)=\sum_{i=1}^{n}(w_i^T Y_i w_i+\lambda_i(w_i^T \mathbf{1}-1))$.

对 w_i 求导有 $\begin{cases}\lambda_i=-\dfrac{2}{\mathbf{1}^T Y_i^{-1}\mathbf{1}}\\ w_i=\dfrac{Y_i^{-1}\mathbf{1}}{\mathbf{1}^T Y_i^{-1}\mathbf{1}}\end{cases}$

$\therefore (w_{i1},\cdots,w_{ik})^T=\dfrac{Y_i^{-1}\mathbf{1}}{\mathbf{1}^T Y_i^{-1}\mathbf{1}}$ (对 k 近邻 j)

对其他 j 有 $w_{ij}=0$

(b) 旋转不变性：$\sum_{i=1}^{n}e_i=\sum_{i=1}^{n}w_i^T Q(x_i-x_j)(x_i-x_j)^T Q^T w_i$

同样 $Y_i=\sum_{s\in S_i}(x_i-x_s)(x_i-x_s)^T$

$L(w,\lambda)=\sum_{i=1}^{n}w_i Q Y_i Q^T w_i+\lambda_i(w_i^T\mathbf{1}-1)$ 且 $w_i\mathbf{1}=1$.

对 w_i 求导得 $w_i=\dfrac{(QY_iQ^T)^{-1}\mathbf{1}}{\mathbf{1}^T(QY_iQ^T)^{-1}\mathbf{1}}$.

$\because QQ^T=I$ $\therefore w_i=\dfrac{Y_i^{-1}\mathbf{1}}{\mathbf{1}^T Y_i\mathbf{1}}$ 同样有旋转不变性.

平移：$\sum_{i=1}^{n}e_i=\sum_{i=1}^{n}\|x_i+t-\sum_{s\in S_i}w_{is}(x_s+t)\|^2=\sum_{i=1}^{n}\|\sum_{s\in S_i}w_{is}(x_i+t-\sum_{s\in S_i}w_{is}(x_s+t))\|^2$

$=\sum_{i=1}^{n}\|\sum_{s\in S_i}w_{is}(x_i-x_s)\|^2$ 和平移前 $\sum_{i=1}^{n}e_i$ 相同.

$\therefore$ 解出的 w_j 一定相同，有平移不变性.

缩放：$\sum_{i=1}^{n}e_i=\sum_{i=1}^{n}\|Sx_i-\sum_{s\in S_i}w_{is}Sx_s\|^2=\sum_{i=1}^{n}\|S\sum_{j=1}^{n}w_{ij}(x_i-x_j)\|^2$

$=S^2\sum_{i=1}^{n}w_i^T(x_i-x_j)(x_i-x_j)^T w_i=S^2\sum_{i=1}^{n}w_i^T Y_i w_i$

$L(w,\lambda)=S\sum_{i=1}^{n}w_i^T Y_i w_i+\lambda_i(w_i\mathbf{1}-1)$ 且 $w_i^T\mathbf{1}-1=0$.

求导得 $w_i=\dfrac{Y_i^{-1}\mathbf{1}}{\mathbf{1}^T Y_i^{-1}\mathbf{1}}$ $\therefore$ 有缩放不变性.

(c) ∵ $\|x_i - \sum_{s\in S_i} w_{is} x_s\|^2$ 尽可能小使得 x_i 可以用 k 近邻样本近似表示、

∴ w_{ij} 保存了 x_i 和其 k 近邻样本之间的几何关系 而 x_i 映射到 d 维空间时，∵ $\|y_i - \sum_{j=1}^{n} w_{ij} y_j\|^2$ 尽量小使得 y_i 和对应的 y_j (k 近邻) 之间也有线性表示 ∴ 保留了局部几何关系、

第一个约束去中心化，第二个约束防止 y_i 越小函数就会一直变小，规范数据。

平移、缩放被消去。

平移不消去会不在中心，缩放会影响数据大小。

(d) $W_i = (w_{i1} \cdots w_{in})^T$

∴ $\sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} w_{ij} y_j\|^2$ ①

$= \sum_{i=1}^{n}\sum_{j=1}^{n} W_i^T (y_i - y_j)(y_i - y_j)^T W_j$

∵ $\sum_{i=1}^{n} y_i y_i^T = I$ 且 $M = (I-W)^T(I-W)$

∴ ① $= \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_i^T y_j$

(e) $\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_i^T y_j = tr(YMY^T) \\ \sum_{i=1}^{n} y_i y_i^T = I \end{cases}$

令 $L(Y, \lambda) = tr(YMY^T + \lambda(YY^T - I))$

求导可知 $M = (Y^T)^{-1} \Lambda Y^T$ ∵ Λ 中 $\lambda_i \geq 0$ ∴ M 是半正定的

$YY^T = I$ 知 Y 可逆

∵ $M \cdot 1 = (I-W)^T (I-W) \cdot 1$

$= (I-W)^T (I \cdot 1 - W \cdot 1) = 0$

∴ $\lambda = 0$ 特征向量为 1