Homework #3

Due: 2024-11-17 23:59 | 7 Problems, 100 Pts

Name: 徐靖, ID: 2200012917

Problem 1 (10'). Suppose that the singular value decomposition of square matrix $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top$. Find out the maximum value of $\|\boldsymbol{A} - \boldsymbol{W}\|_F$ where $\boldsymbol{W}$ is an orthogonal matrix. ◀

Answer. $\forall W$, let $W' = -U^\top W V$. Then W is an orthogonal matrix, which is equivalent to W' being an orthogonal matrix. And we have

$$\|A - W\|_F = \|U\Sigma V^\top + UW'V\|_F = \|\Sigma + W'\|_F = \|\Sigma W'^\top + I\|_F$$

Assume $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\}, W' = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix}, w_i = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix}^\top$. Let $\delta_{ij} = 1(i=j)$ or $-1(i \neq j)$ is Kronecker delta. Then we have

$$\begin{aligned}\|A - W\|_F &= \sum_{i=1}^n \sum_{j=1}^n (\sigma_i a_{ij} + \delta_{ij})^2 \\ &= n + \sum_{i=1}^n \sigma_i^2 + 2\sum_{i=1}^n \sigma_i a_{ii}\end{aligned}$$

Given that $a_{ii} \leq \sqrt{|w_i|} = 1$, thus $\|A - W\|_F \leq n + \sum_{i=1}^n \sigma_i^2 + 2\sum_{i=1}^n \sigma_i$. The inequality is equal when $W' = I$

In conclusion,

$$\max \|A - W\|_F = n + \mathrm{tr}(\Sigma^2 + 2\Sigma)$$

◁

Problem 2 (12'). Suppose $\boldsymbol{A} = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ is the SVD of matrix $\boldsymbol{A}$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $r \geq 1$. Answer the following problems. You don't need to prove your result.

(1) (3') Determine whether there exists $\alpha \in \mathbb{R}$, such that there exists an absolute constant $c > 0$, for any $r \geq 1$, any matrix $\boldsymbol{A}$ and $k = 1, 2, \cdots, r$,

$$\min_{\mathrm{rank}(\boldsymbol{B}) \leq k} \frac{\|\boldsymbol{A} - \boldsymbol{B}\|_F}{\|\boldsymbol{A}\|_F} \leq ck^\alpha.$$

If so, write down the value of minimum α as well.

(2) (3') Determine whether there exists $\alpha \in \mathbb{R}$, such that there exists an absolute constant $c > 0$, for any $r \geq 1$, any matrix $\boldsymbol{A}$ and $k = 1, 2, \cdots, r$,

$$\min_{\mathrm{rank}(\boldsymbol{B}) \leq k} \frac{\|\boldsymbol{A} - \boldsymbol{B}\|_F}{\|\boldsymbol{A}\|_2} \leq ck^\alpha.$$

If so, write down the value of minimum α as well.

(3) (3') Determine whether there exists $\alpha \in \mathbb{R}$, such that there exists an absolute constant $c > 0$, for any $r \geq 1$, any matrix $\boldsymbol{A}$ and $k = 1, 2, \cdots, r$,

$$\min_{\operatorname{rank}(\boldsymbol{B}) \leq k} \frac{\|\boldsymbol{A} - \boldsymbol{B}\|_2}{\|\boldsymbol{A}\|_F} \leq c k^{\alpha}.$$

If so, write down the value of minimum α as well.

(4) (3') Determine whether there exists $\alpha \in \mathbb{R}$, such that there exists an absolute constant $c > 0$, for any $r \geq 1$, any matrix $\boldsymbol{A}$ and $k = 1, 2, \cdots, r$,

$$\min_{\operatorname{rank}(\boldsymbol{B}) \leq k} \frac{\|\boldsymbol{A} - \boldsymbol{B}\|_2}{\|\boldsymbol{A}\|_2} \leq c k^{\alpha}.$$

If so, write down the value of minimum α as well.

◀

Answer. •

(1) $\alpha_{\min} = 0$

(2) $\neg \exists \alpha$

(3) $\alpha_{\min} = -\frac{1}{2}$

(4) $\alpha_{\min} = 0$

◁

Problem 3 (26'). Recall the Johnson-Lindenstrauss lemma we learned in class.

Theorem 1 (Johnson-Lindenstrauss Lemma)**.** *Given $\epsilon \in (0, 1)$ and n vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^m$. Pick a random matrix $\boldsymbol{\Pi} \in \mathbb{R}^{k \times m}$ as $\boldsymbol{\Pi} = \frac{1}{\sqrt{k}} \boldsymbol{W}$ where each entry $W_{i,j}$ $(1 \leq i \leq k, 1 \leq j \leq m)$ is sampled independently from $\mathcal{N}(0, 1)$. Then there exists an absolute constant $c_1 > 0$, such that when $k \geq \frac{c_1 \ln n}{\epsilon^2}$, with probability at least $1 - \frac{3}{2n}$, for all $i, j \in \{1, 2, \cdots, n\}$,*

$$(1 - \epsilon) \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\boldsymbol{\Pi} \boldsymbol{x}_i - \boldsymbol{\Pi} \boldsymbol{x}_j\|_2 \leq (1 + \epsilon) \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2.$$

In this problem, we are going to use Johnson-Lindenstrauss lemma to prove all big matrices are approximately low-rank.

(1) (2') Given $\epsilon \in (0, 1)$ and n vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^m$. Prove that, there exists an absolute constant $c_2 > 0$, such that when $k \geq \frac{c_2 \ln n}{\epsilon^2}$, there exists a matrix $\boldsymbol{\Pi} \in \mathbb{R}^{k \times m}$ satisfying

$$(1 - \epsilon) \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 \leq \|\boldsymbol{\Pi} \boldsymbol{x}_i - \boldsymbol{\Pi} \boldsymbol{x}_j\|_2^2 \leq (1 + \epsilon) \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$$

for all $i, j \in \{1, 2, \cdots, n\}$.

(2) (7') Given $\epsilon \in (0,1)$ and n vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^m$. Prove that, there exists an absolute constant $c_3 > 0$, such that for $r \geq \frac{c_3 \ln(n+1)}{\epsilon^2}$, there exists a matrix $\boldsymbol{Q} \in \mathbb{R}^{r \times m}$ satisfying

$$|\boldsymbol{x}_i^\top \boldsymbol{x}_j - \boldsymbol{x}_i^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{x}_j| \leq \epsilon(\|\boldsymbol{x}_i\|_2^2 + \|\boldsymbol{x}_j\|_2^2 - \boldsymbol{x}_i^\top \boldsymbol{x}_j)$$

for all $i, j \in \{1, 2, \cdots, n\}$.

[Hint: Consider $2\boldsymbol{x}_i^\top \boldsymbol{x}_j = \|\boldsymbol{x}_i\|_2^2 + \|\boldsymbol{x}_j\|_2^2 - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$.]

(3) (7') For any matrix $\boldsymbol{M}$, define

$$\|\boldsymbol{M}\|_{\max} = \max_{i,j} |M_{ij}|.$$

Given matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$.

(a) (2') Prove that, $\|\boldsymbol{X}\|_{\max} \leq \|\boldsymbol{X}\|_2$.

(b) (5') Prove that, there exists matrix $\boldsymbol{U} = (\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_m) \in \mathbb{R}^{n \times m}, \boldsymbol{V} = (\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_n) \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{X} = \boldsymbol{U}^\top \boldsymbol{V}$ and $\|\boldsymbol{u}_i\|_2^2 \leq \|\boldsymbol{X}\|_2, \ \|\boldsymbol{v}_j\|_2^2 \leq \|\boldsymbol{X}\|_2$.

(4) (5') Suppose $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ where $m \geq n$ and $\epsilon \in (0,1)$. Prove that, there exists an absolute constant $c_4 > 0$, such that with $r = \left\lceil \frac{c_4 \ln(m+n+1)}{\epsilon^2} \right\rceil$,

$$\min_{\operatorname{rank}(\boldsymbol{Y}) \leq r} \|\boldsymbol{X} - \boldsymbol{Y}\|_{\max} \leq \epsilon \|\boldsymbol{X}\|_2.$$

(5) (5') A side note is that the result in problem (4) doesn't work for small matrix (small n). Consider $\boldsymbol{X} = \boldsymbol{I}_2$. Find out the value of

$$\min_{\operatorname{rank}(\boldsymbol{Y}) \leq 1} \|\boldsymbol{X} - \boldsymbol{Y}\|_{\max}.$$

Prove your result.

◀

Answer. (1) Note that

$$\|\Pi x\| \leq (1+\epsilon)\|x\| \Rightarrow \|\Pi x\|^2 \leq (1+3\epsilon)\|x\|$$

So we have $k \geq \frac{c_1 \ln n}{9\epsilon^2}$, let $c_2 = \frac{c_1}{9}$ then the proof is complete.

(2) In fact,

$$\begin{aligned}
|x_i^\top x_j - x_i^\top Q^\top Q x_j| &= \frac{1}{2}|\|x_i\|^2 + \|x_j\|^2 - \|x_i - x_j\|^2 - \|Qx_i\|^2 - \|Qx_j\|^2 + \|Qx_i - Qx_j\|^2| \\
&\leq \frac{1}{2}|\|Qx_i\|^2 - \|x_i\|^2| + \frac{1}{2}|\|Qx_j\|^2 - \|x_j\|^2| + \frac{1}{2}|\|Qx_i - Qx_j\|^2 - \|x_i - x_j\|^2| \\
&\leq \frac{\epsilon}{2}(\|x_i\|^2 + \|x_j\|^2 + \|x_i - x_j\|^2) \\
&= \epsilon(\|x_i\|^2 + \|x_j\|^2 - x_i^\top x_j)
\end{aligned}$$

(3) (a) Note that,

$$\|X\|_2 = \max_{\|a\|_2=1} \|Xa\|$$

If $|X_{ij}| = \|X\|_{\max} > \|X\|_2$, then let $a = e_j$, we have $\|Xa\|_2 = \sqrt{\sum_{i\in m} X_{ij}^2} \geq |X_{ij}| > \|X\|_2$. This contradicts.

(b) $X = U_X \Sigma V_X^\top$. Let $U = \Sigma^{\frac{1}{2}} U_X^\top, \Sigma^{\frac{1}{2}} V_X^\top$. Then $X = U^\top V$, $\|u_i\|_2^2 = \|Ue_i\|_2^2 = (U_x^\top e_i)^\top \Sigma (U_x^\top e_i) \leq \sigma_1$ for $\|U_i^\top e_i\|_2 = 1$. Similarly $\|v_j\|_2^2 \leq \sigma_1$

(4) By using 3.(b), $X = U^\top V, X_{ij} = U_i^\top V_j$. Let Q be the random projection matrix. $\tilde{u_i} = Qu_i, \tilde{v_i} = Qv_i, Y_{ij} = \tilde{u_i}^\top \tilde{v_k}$ Then there are $m+n$ vectors $(u_1, \dots, u_m, v_i, \dots, v_n)$, and we have

$$|X_{ij} - Y_{ij}| \leq \epsilon(\|u_i\|^2 + \|v_j\|^2 - u_i^\top v_j) \leq \epsilon(\|u_i\|^2 + \|v_j\|^2 + \|u_i\|\|v_j\|) \leq 3\epsilon \|X\|_2$$

Proof completed.

(5) Let, $y = (a, b)^\top$, then we have

$$Y = yy^\top = \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}$$

And

$$\|X - Y\|_{\max} = \left\| \begin{pmatrix} 1-a^2 & ab \\ ab & 1-b^2 \end{pmatrix} \right\|_{\max} = \max |1-a^2|, |1-b^2|, |ab|$$

Thus $\min \|X - Y\|_{\max} = \frac{1}{2}$

◁

Problem 4 (22').

(1) (8') Given $\boldsymbol{S} = \mathrm{diag}(s_1, s_2, \cdots, s_n)$ where $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$. Find out the value of

$$\max_{\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_r} \|\boldsymbol{W}^\top \boldsymbol{S}\|_F$$

where $\boldsymbol{W} \in \mathbb{R}^{n\times r}$ $(n \geq r)$.

(2) (14') Suppose the singular values of a $m \times n$ $(m \geq n)$ matrix $\boldsymbol{A}$ are $\sigma_i(\boldsymbol{A})$, and $\sigma_1(\boldsymbol{A}) \geq \cdots \geq \sigma_n(\boldsymbol{A})$.

(a) (12') Prove that, for any $k \in [n]$,

$$\sum_{i=1}^{k} \sigma_i(\boldsymbol{A}) = \max_{\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_k,\, \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_k} |\mathrm{Tr}(\boldsymbol{U}^\top \boldsymbol{A} \boldsymbol{V})|,$$

where $\boldsymbol{I}_k$ is the rank-k identity matrix, $\boldsymbol{U}$ is a $m \times k$ matrix, and $\boldsymbol{V}$ is a $n \times k$ matrix.

(b) (2') Prove that, for any $k \in [n]$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ $(m \geq n)$, $\sum_{i=1}^{k} \sigma_i(\boldsymbol{A}+\boldsymbol{B}) \leq \sum_{i=1}^{k} \sigma_i(\boldsymbol{A}) + \sum_{i=1}^{k} \sigma_i(\boldsymbol{B})$.

◀

Answer. (1) $\max_W \|W^\top S\|_F = \sqrt{s_1^2+s_2^2+\cdots+s_r^2}$. This upper bound can be reached with $W = (w_1, w_2, \ldots, w_r)$, where w_i is the column vector with 1 on the i-th dimension and 0 on the rest. The proof of $\max_W \|W^\top S\|_F \le \sqrt{s_1^2+s_2^2+\cdots+s_r^2}$ is as follows.

$\forall W$, construct orthgonal matrix $U \in \mathbb{R}^{n\times n}$ such that the first r column of U is the same to W.

Consider matrix R which satisfies $R_{ij} = U_{ij}^2$. Since U is orthogonal, the row sums and column sums of U are all 1

Apply Birkhoff's Theorem on R, then we have $R = \sum_{i=1}^k \theta_i P_i$, where $\sum_{i=1}^k \theta_i = 1$

$\forall$ permutation matrix P, suppose that $P_{i,q_i} = 1$ (then q_i is a permutaion of $1, \ldots, n$), we have $\sum_{i=1}^r \sum_{j=1}^n s_j^2 P_{ij} = \sum_{i=1}^r s_{q_i}^2 \le \sum_{i=1}^r s_i^2$ Thus,

$$\begin{aligned}
\|W^\top S\|_F^2 &= \sum_{i=1}^r \sum_{j=1}^n s_j^2 W_{ij}^2 = \sum_{i=1}^r \sum_{j=1}^n s_j^2 R_{ij} = \sum_{i=1}^r \sum_{j=1}^n s_j^2 (\sum_{l=1}^k \theta_l (P_i)_{ij}) \\
&= \sum_{l=1}^k \theta_l \sum_{i=1}^r \sum_{j=1}^n s_j^2 (P_i)_{ij} \le \sum_{l=1}^k \theta_l \sum_{i=1}^r s_i^2 = \sum_{i=1}^r s_i^2
\end{aligned}$$

(2) (a) Consider the SVD of A as $A = U_0 D V_0$, where $U_0 \in \mathbb{R}^{m\times m}, D \in \mathbb{R}^{m\times n}, V_0 \in \mathbb{R}^{n\times n}$ and $U_0^\top U_0 = I_m, V_0^\top V_0 = I_n, D_{ii} = \sigma_i$.

For all k and $U \in \mathbb{R}^{m\times k}, V \in \mathbb{R}^{n\times k}$ that $U^\top U = V^\top V = I_k$, we have $\left(U_0^\top U\right)^\top \left(U_0^\top U\right) = U^\top U_0 U_0^\top U = U^\top I_m U = I_k$, as well as $(V_0 V)^\top (V_0 V) = I_k$.

We rename these two matrix as $L = U_0^\top U$ and $R = V_0 V$. As $L^\top L = R^\top R = I_k$, we can see that all the column vectors of L and R are unit vectors, as well as all the row vectors are not longer than 1. (since you can always extend L and R to a orthogonal matrix, afterward the length of all the row vectors will be 1.)

$$\begin{aligned}
\operatorname{tr}\left(U^\top A V\right) &= \operatorname{tr}\left(U^\top U_0 D V_0 V\right) = \operatorname{tr}\left(\left(U_0^\top U\right)^\top D V_0 V\right) = \operatorname{tr}\left(L^\top D R\right) \\
&= \sum_{i=1}^n \sum_{j=1}^k L_{ji}^\top \sigma_i R_{ij} = \sum_{i=1}^n \sigma_i \sum_{j=1}^k L_{ij} R_{ij}
\end{aligned}$$

Let $s_i = \sum_{j=1}^k L_{ij} R_{ij}$, and we have the following constraints on s :

* for all $i, s_i \le 1$. This is because that $s_i = \sum_{j=1}^k L_{ij} R_{ij}$ is the inner product of two row vectors whose length is not more than 1.
* for all $j, \sum_{j=1}^i s_j \le k$. This is because that $\sum_{j=1}^i s_j = \sum_{j=1}^i \sum_{l=1}^k L_{jl} R_{jl} = \sum_{l=1}^k \sum_{j=1}^i L_{jl} R_{jl}$ is the sum of inner products on the first i dimensions of k pairs of column vectors whose length is not more than 1.

Using Abel transformation (assume that $\sigma_{n+1} = 0$), we have: $\operatorname{tr}\left(U^\top A V\right) = \sum_{i=1}^n \sigma_i s_i = \sum_{i=1}^n \left(\sigma_i - \sigma_{i+1}\right) t_i$, where $t_i = \sum_{j=1}^i s_j \le \min\{i, k\}$ (because $s_i \le 1$ and $\sum_{j=1}^i s_j \le k$) As $\sigma_i - \sigma_{i+1} \ge 0$ holds for all i :

$$\operatorname{tr}\left(U^\top A V\right) = \sum_{i=1}^n \left(\sigma_i - \sigma_{i+1}\right) t_i \le \sum_{i=1}^k \left(\sigma_i - \sigma_{i+1}\right) i + \sum_{i=k+1}^n \left(\sigma_i - \sigma_{i+1}\right) k = \sum_{i=1}^k \sigma_i$$

Symmetrically we can prove that $\operatorname{tr}\left(U^\top AV\right) \geq -\sum_{i=1}^n \sigma_i$, thus $\left|\operatorname{tr}\left(U^\top AV\right)\right| \leq \sum_{i=1}^n \sigma_i$. Construct $U^* = U_0^\top I_{mk}$ and $V^* = V_0^\top I_{nk}$ (where $I_{ij} \in \mathbb{R}^{i\times j}$ having only the main diagonal as 1), then we have $\operatorname{tr}\left(U^{*\top} AV^*\right) = \operatorname{tr}\left(I_{km} D I_{nk}\right) = \sum_{i=1}^k \sigma_i$. So $\max\left|\operatorname{tr}\left(U^\top AV\right)\right| = \sum_{i=1}^k \sigma_i$.

(b) As a conclusion of 4.2(b), we instantly have

$$
\begin{aligned}
\sum_{i=1}^k \sigma_i(A+B) &= \max_{U^\top U = V^\top V = I_k} \left|\operatorname{tr}\left(U^\top (A+B) V\right)\right| = \left|\operatorname{tr}\left(U^{*\top}(A+B)V^*\right)\right| \\
&= \left|\operatorname{tr}\left(U^{*\top} AV^*\right) + \operatorname{tr}\left(U^{*\top} BV^*\right)\right| \leq \left|\operatorname{tr}\left(U^{*\top} AV^*\right)\right| + \left|\operatorname{tr}\left(U^{*\top} BV^*\right)\right| \\
&\leq \max_{U^\top U = V^\top V = I_k} \left|\operatorname{tr}\left(U^\top AV\right)\right| + \max_{U^\top U = V^\top V = I_k} \left|\operatorname{tr}\left(U^\top BV\right)\right| \\
&= \sum_{i=1}^k \sigma_i(A) + \sum_{i=1}^k \sigma_i(B).
\end{aligned}
$$

◁

Problem 5 (5'). Use power method presented in section 3.7.1 of textbook to compute the largest 10 singular values of the random matrix $\boldsymbol{A}$ generated using NumPy as follows.

Algorithm 1: Generate the matrix

```
np.random.seed(20241025)      // Set random seed.
A = np.random.randn(2000, 1000)
```

Write down the singular values and submit the code as an attachment. You could use any functions provided by the NumPy package. If you use other programming languages or other packages to solve this problem, please generate the random matrix in a similar fashion. ◀

Answer. 75.72 75.54 75.4 74.95 74.74 74.39 74.25 74.21 73.91 73.85 ◁

Problem 6 (5'). Consider a labeling $f : \{0,1\}^d \to \{-1,+1\}$: $f(x_1, x_2, x_3, \cdots, x_d) := (-1)^{\sum_{i=1}^d x_i}$. Is $\{0,1\}^d$ linearly separable if it is labeled by f? Prove your result. ◀

Answer. Obviously, it is not linearly separable. If it is linearly separable, then $\{0,1\}^2$ on the plane $x_3 = x_4 = \cdots = x_n = 0$ is separated by the straight line projected by the hyperplane on the plane, and among these four points, $(0,0), (1,1)$ has the same label, and $(0,1), (1,0)$ has the opposite label to the first two. Therefore, these four points are not linearly separable. This contradicts the assumption. ◁

Problem 7 (20'). Consider function $K(\boldsymbol{x}, \boldsymbol{y}) = (1 + a\boldsymbol{x}^\top \boldsymbol{y})^2$ where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$ is a constant.

(1) (6') Find out all possible $a \in \mathbb{R}$, such that $K(\boldsymbol{x}, \boldsymbol{y})$ is a kernel function. Prove your result.

(2) (14') Suppose $a \neq 0$ and $K(\boldsymbol{x}, \boldsymbol{y})$ is a kernel function, i.e., there exists $\boldsymbol{\varphi} : \mathbb{R}^n \to \mathbb{R}^m$ such that $K(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\varphi}(\boldsymbol{x}), \boldsymbol{\varphi}(\boldsymbol{y}) \rangle$. Prove that, for any $\boldsymbol{\varphi}$ satisfying the conditions above, $m \geq \frac{n(n+3)}{2} + 1$.

[Hint: Consider matrix $M_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ $(1 \le i, j \le r)$, then $\boldsymbol{M}$ is a semi-definite matrix and $\boldsymbol{M} = \boldsymbol{G}^\top \boldsymbol{G}$ where $\boldsymbol{G} = (\boldsymbol{\varphi}(\boldsymbol{x}_1) \mid \cdots \mid \boldsymbol{\varphi}(\boldsymbol{x}_r))$. Take specific $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r \in \mathbb{R}^n$ and consider the rank of $\boldsymbol{M}$. You may find the basic results in linear algebra that $\mathrm{rk}(\boldsymbol{G}^\top \boldsymbol{G}) = \mathrm{rk}(\boldsymbol{G})$ useful.]

◀

Answer.

(1) $\forall a > 0$, let $\varphi(\boldsymbol{x}) = (1, \sqrt{2a}x_1, \ldots, \sqrt{2a}x_n, ax_1^2, \ldots, ax_n^2, \sqrt{2}ax_1x_2, \ldots, \sqrt{2}ax_1x_n, \ldots, \sqrt{2}ax_{n-1}x_n)$ Then we have

$$\varphi(\boldsymbol{x}) \cdot \varphi(\boldsymbol{y}) = 1 + \sum_{i=1}^{n}(2ax_iy_i + a^2x_i^2y_i^2) + \sum_{1 \le i < j \le n} 2a^2x_iy_ix_jy_j = K(\boldsymbol{x}, \boldsymbol{y})$$

In this case, K is the kernel function.

For $a < 0$,let $x_1 = (1, 0, \ldots, , 0), x_2 = (0, 1, 0, \ldots, 0)$, then we get kernel matrix based on $\{x_1, x_2\}$:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

It is not semi-definite. Thus K is not a kernel function. The case $a = 0$ is trivial.

(2) Construct the following vectors: $a_0 = (0, \ldots, 0)$, $a_{i+} = (0, \ldots, 1, \ldots, 0)$ (only the i-th position is 1), $a_{i-} = (0, \ldots, -1, \ldots, 0)$ (only the i-th position is -1), and $a_{(i,j)} = (0, \ldots, 1, \ldots, 1, \ldots, 0)$ (only the i-th and j-th positions are 1, where $i \neq j$; tuples (i, j) and (j, i) are equivalent).

There are $l = \frac{1}{2}n^2 + \frac{3}{2}n + 1$ such vectors. Consider an $l \times l$ matrix M, where $M_{i,j} = K(a_i, a_j)$ (i, j are indices of the l vectors). It is easy to see that M is a real symmetric matrix. Define m_i as the row vector corresponding to the i-th row.

Noting that $m_0 = (1, \ldots, 1)$, the first column of M is all ones. Subtracting m_0 from every row except the first, the first column of M becomes $[1, 0, \ldots, 0]^\top$. Denote the resulting lower-right $(l-1) \times (l-1)$ submatrix as M'. Then:

$$\mathrm{rk}(M) = \mathrm{rk}(M') + 1,$$

where $M'_{i,j} = K(a_i, a_j) - 1$.

Next, consider the vector $M'_{(i,j)} - M'_i - M'_j$. Let $f(x)$ denote the x-th entry of this vector. For $k, h \in [n] \setminus \{i, j\}$, we have:

$$\begin{aligned}
f(k^+) = f(k^-) &= (1-1) - (1-1) - (1-1) = 0, \\
f(i^+) &= (4-1) - (4-1) - (1-1) = 0, \\
f(i^-) &= (0-1) - (0-1) - (1-1) = 0, \\
f((k,l)) &= (1-1) - (1-1) - (1-1) = 0, \\
f((i,k)) &= (4-1) - (4-1) - (1-1) = 0, \\
f((i,j)) &= (9-1) - (4-1) - (4-1) = 2.
\end{aligned}$$

In other words, $M'_{(i,j)} - M'_i - M'_j$ is zero everywhere except at the (i,j)-th position, where it equals 2. Thus, removing the (i,j)-th row and column from M' reduces its rank by exactly 1.

Now, consider the top-left $2n \times 2n$ submatrix M'' of M'. After simple row and column exchanges, we observe:

$$|M''| = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}^n > 0.$$

Therefore, M'' has full rank, and M is full rank.

Finally, consider a matrix $G = \langle \varphi(a_0) \mid \cdots \mid \varphi(a_{(n-1,n)}) \rangle$, with G having l columns and m rows. It is easy to verify that $M = G^\top G$. Consequently:

$$\mathrm{rk}(G) = \mathrm{rk}(G^\top G) = \mathrm{rk}(M) = l.$$

Thus:

$$m \geq \mathrm{rk}(G) = \frac{1}{2}n^2 + \frac{3}{2}n + 1.$$

$\triangleleft$