Exercise 01 of Machine Learning [IN 2064]

Name: Yiman Li
Matr-Nr: 03724352
cooperate with Kejia Chen(03729686)

Problem 1

In order for the matrix product to exist, the number of columns in the former matrix must equal the number of rows in the latter matrix, here the function f can be described as:

$$f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) = (\boldsymbol{x}^T)^{1\times M}\boldsymbol{A}y^{N\times 1} + \boldsymbol{B}\boldsymbol{x}^{M\times 1} - (\boldsymbol{y}^T)^{1\times N}\boldsymbol{C}\boldsymbol{Z}^{P\times Q}\boldsymbol{D} - (\boldsymbol{y}^T)^{1\times N}\boldsymbol{E}^T\boldsymbol{y}^{N\times 1} + \boldsymbol{F}$$

Then using the aforementioned rules, we can get

$$\boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{B} \in \mathbb{R}^{1\times M}, \boldsymbol{C} \in \mathbb{R}^{N\times P}, \boldsymbol{D} \in \mathbb{R}^{Q\times 1}, \boldsymbol{E} \in \mathbb{R}^{N\times N}, \boldsymbol{F} \in \mathbb{R}^{1\times 1}$$

Problem 2

Now that the function $f(x) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j M_{ij} \in \mathbb{R}^{N\times N}$, so we can simply rewrite the function as $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}$ just on the condition that the number of columns in the former matrix must equal the number of rows in the latter matrix in matrix product.

Problem 3

a) When $R(\boldsymbol{A}) = R(\boldsymbol{A}, \boldsymbol{b}) = M$, then we can safely say that the solution is unique for each choice of b. Here R represents "Rank of the Matrix".
b) Notice that the matrix has a zero eigenvalue, which means the $R(\boldsymbol{A} < 5)$, so we can not safely draw a conclusion that this matrix is diagonalizable, so the Equation (1) can not always find a unique solution $\boldsymbol{x}$ for any choice of $\boldsymbol{b}$.

Problem 4

Here we can use the property that the determinant of $\boldsymbol{A}$ is equal to the product of its eigenvalues $|\boldsymbol{A}| = \prod_{i=1}^{n}\lambda_i$. Now that the matrix $\boldsymbol{A}$ is invertible, which means that $|\boldsymbol{A}| \neq 0$, so we can say that one of the eigenvalues of matrix $\boldsymbol{A}$ is zero since $\boldsymbol{A}$ is invertible.

Problem 5

According to reference [**?**], since a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ has orthogonal eigenvectors and is thus orthogonal, we can therefore represent $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$, then we can show that

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{U}\Lambda\boldsymbol{U}^T\boldsymbol{x} = \boldsymbol{y}^T\Lambda\boldsymbol{y} = \sum_{i=1}^{n}\lambda_i y_i^2$$

where $\boldsymbol{y} = \boldsymbol{U}^T\boldsymbol{x}$. Because y_i^2 is always positive, the sign of this expression depends entirely on the $\lambda_i's$. So if all eigenvalues $\lambda_i's \geq 0$, then it is positive semidefinite.

Conversely, if a matrix is positive semidefinite, then we assume that this matrix $\boldsymbol{A}$ has a negative eigenvalue $\lambda < 0$ and its correspongding eigenvector $\boldsymbol{x}$, so we can get

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$

then we multiply the two sides of the equetion by the transpose of the eigenvector $\boldsymbol{x}^T$, which means

$$\boldsymbol{x}^T \boldsymbol{Ax} = \boldsymbol{x}^T \lambda \boldsymbol{x} = \lambda \boldsymbol{x}^T \boldsymbol{x} = \lambda \boldsymbol{x}^2 < 0$$

this result is contrary to the condition that the matrix is positive semidefiniten, so the assumption that the matrix has a negative eigenvalue doesn't make sense. So in the end we can draw a conclusion that a positive semidefinite matrix has no negative eigenvalues.

Problem 6

Choose an arbitrary vextor $\boldsymbol{x} \in \mathbb{R}^N$ for the matrix $\boldsymbol{B}$, the scalar value can be written as

$$\boldsymbol{x}^T \boldsymbol{Bx} = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{Ax} = (\boldsymbol{Ax})^T (\boldsymbol{Ax}) = (\boldsymbol{Ax})^2 \geq 0$$

So the matrix $\boldsymbol{B}$ is positive semi-definite for any choice of $\boldsymbol{A}$.

Problem 7

a)

1. Using the second partial derivatives of function f, that is
$$\mathrm{d}_x^2 f(x) = \mathrm{d}_x(ax + b) = a$$
So when $a > 0$, then f(x) becomes a strictly convex quadratic function, which has at most one global minimum.

2. When $a = b = 0$, then f(x) is a constant function, so each point of x is the solution of the minimum value of f(x).

3. When $a < 0$, then f(x) becomes a strictly concave function; or when $a = 0, b \neq 0$, whose minimum value lies on infinity. In both situation, there is no solution for the optimization.

b) Now we try to equal the first partial derivatives of function f to zero, that is

$$\mathrm{d}_x f(x) = ax + b = 0$$

So the point $x = -\frac{b}{a}$ minimizes the objective function.

Problem 8

a) Now that the matrix is symmetric and positive simidefinite, so the Hessian $\nabla_x^2 g(\boldsymbol{x})$ of the objective function is shown as below:

$$\nabla_x^2 g(\boldsymbol{x}) = \nabla_x^2 [(\frac{1}{2})\boldsymbol{x}^T \boldsymbol{Ax} + \boldsymbol{bx}^T + c] = \boldsymbol{A}$$

Similar to $\textbf{Problem7}$, when $\boldsymbol{A} > 0$, which means that the matrix is positive definite, then the optimization problem becomes a strictly convex function(See definition in reference [**?**])

b) If $\boldsymbol{A}$ is positive definite, then we can find a unique solution for the optimization at the point where the gradient of the function is zero; if $\boldsymbol{A}$ is positive semidefinite, then we might find the solution at infinity. However, when $\boldsymbol{A}$ has a negative eigenvalue, which means that the matrix is indefinite, the function $g(x)$ may even become a concave function, so we may not find the solution for the optimization.

c) Now that $\boldsymbol{A}$ is positive definite(PD), so we just try to equal the gradient of the function to zero to figure out the result, that is:

$$\nabla_x g(\boldsymbol{x}) = \nabla_x [(\frac{1}{2})\boldsymbol{x}^T \boldsymbol{Ax} + \boldsymbol{bx}^T + c] = \boldsymbol{Ax} + \boldsymbol{b} = 0$$

So we can get to the point that when the matrix $\boldsymbol{A}$ is positive definite (PD), then this optimization problem $g(\boldsymbol{x})$ will have a unique solution at the point $\boldsymbol{x} = -\boldsymbol{A}^{-1}\boldsymbol{b}$.

Problem 9

According to

$$p(A|B,C) = \frac{p(A,B,C)}{p(B,C)} = \frac{\frac{p(A,B,C)}{p(C)}}{\frac{p(B,C)}{p(C)}} = \frac{p(A,B|C)}{p(B|C)} = p(A|C)$$

that is

$$p(A,B|C) = p(A|C)p(B|C)$$

so we can say that two events A and B are conditionally independent given an event C with $P(C) > 0$. However, according to reference [**?**], conditional independence cannot lead to the concept of independence. Here is an example. Assuming that a box contains two coins: a regular coin and one irregular coin with $(P(H) = 1)$. Now choose a coin at random and toss it twice. Define the following events.

- A = First coin toss results in an H.
- B = Second coin toss results in an H.
- C = The regular coin has been chosen.

From this example we can get that $p(A|B,C) = p(A|C) = \frac{1}{2}$, then we can calculate that

$$p(A,B) = p(A,B|C) + p(A,B|\overline{C}) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + 1\cdot 1\cdot\frac{1}{2} = \frac{5}{8}$$

whereas

$$p(A) = p(A|C) + p(A|\overline{C}) = \frac{1}{2}\cdot\frac{1}{2} + 1\cdot\frac{1}{2} = \frac{3}{4}$$

$$p(B) = p(B|C) + p(B|\overline{C}) = \frac{1}{2}\cdot\frac{1}{2} + 1\cdot\frac{1}{2} = \frac{3}{4}$$

On this condition $p(A,B) \neq p(A)p(B)$, so this statement is wrong.

Problem 10

When $p(A|B,C) = p(A|C)$, we can get the information that A and B are conditionally independent when given C. But generally speaking, conditional independence neither implies (nor is it implied by) independence, which is the information in $p(A|B) = p(A)$. Consider rolling a die and let

$$A = \{1,2\}, B = \{2,4,5\}, C = \{1,4\}$$

so we can get

$$p(A) = \frac{1}{3}, p(B) = \frac{1}{2}, p(A,B) = \frac{1}{6} = p(A)p(B)$$

which means A and B are independent. But we can also figure that

$$p(A|B,C) = \frac{p(\{1,2\})}{p(\{4\})} = 0, p(A|C)\frac{p(\{1,2\})}{p(\{1,4\})} = \frac{1}{2}$$

so the statement is false.

Problem 11

This problem is based on the concept of probability density function, more details can be found in reference [**?**]. According to the concept, we can directly write down the corresponding formular as below:
(1)

$$p(a) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(a,b,c)\mathrm{d}b\mathrm{d}c$$

(2)

$$p(c|a,b) = \frac{p(a,b,c)}{p(a,b)} = \frac{p(a,b,c)}{\int_{-\infty}^{\infty} p(a,b,c)\mathrm{d}c}$$

(3)

$$p(b|c) = \frac{p(b,c)}{p(c)} = \frac{\int_{-\infty}^{\infty} p(a,b,c)\mathrm{d}a}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(a,b,c)\mathrm{d}a\mathrm{d}b}$$

Problem 12

The probability that a person has a positive test result is

$$\frac{1}{1000} \times 0.95 + \frac{999}{1000} \times 0.05 = \frac{5090}{100000}$$

So when man obtains a positive result, his/her probability of having the disease is

$$\frac{\frac{1}{1000} \times 0.95}{\frac{1}{1000} \times 0.95 + \frac{999}{1000} \times 0.05} = \frac{19}{1018} = 0.0187$$

Problem 13

Since the mean value of a Gaussian distribution is μ, using the property that $Var[X] = E[X^2] + E[X]^2$, we can easily figure out that $E[f(x)] = a\mu + b(\mu^2 + \sigma^2) + c$

Problem 14

Assume $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \sum)$, according to reference **?** , we have the following conclusion:

$$\begin{aligned} E[\boldsymbol{Ax}] &= \boldsymbol{A}E[\boldsymbol{x}] & (1) \\ E(\boldsymbol{xx}^T) &= \sum + \boldsymbol{mm}^T & (2) \\ Var[\boldsymbol{Ax}] &= \boldsymbol{A}Var[\boldsymbol{x}]\boldsymbol{A}^T & (3) \\ E[\boldsymbol{x}^T\boldsymbol{Ax}] &= Tr(\boldsymbol{A}\sum) + \boldsymbol{m}^T\boldsymbol{Am} & (4) \end{aligned}$$

- Using the Equation 1, we can get that

$$E[g(\boldsymbol{x})] = E[\boldsymbol{Ax}] = \boldsymbol{A}E[\boldsymbol{x}] = \boldsymbol{A\mu}$$

- Since $g(\boldsymbol{x}) = \boldsymbol{Ax} \sim \mathcal{N}(\boldsymbol{A\mu}, \boldsymbol{A}\sum\boldsymbol{A}^T)$, so according to the Equation 2, we can get

$$E[g(\boldsymbol{x})g(\boldsymbol{x})^T] = E[\boldsymbol{Axx}^T\boldsymbol{A}^T] = \boldsymbol{A}E[\boldsymbol{xx}^T]\boldsymbol{A}^T = \boldsymbol{A}(\sum + \boldsymbol{\mu\mu}^T)\boldsymbol{A}^T$$

- Since $g(\boldsymbol{x}) = \boldsymbol{Ax} \sim \mathcal{N}(\boldsymbol{A\mu}, \boldsymbol{A}\sum\boldsymbol{A}^T)$, the changing the Equation 4 by replacing $\boldsymbol{A}$ by an Identity matrix $\boldsymbol{I}$, we can get the result as below:

$$E[g(\boldsymbol{x})^T g(\boldsymbol{x})] = E[g(\boldsymbol{x})^T\boldsymbol{I}g(\boldsymbol{x})] = Tr(\boldsymbol{A}\sum\boldsymbol{A}^T) + \boldsymbol{\mu}^T\boldsymbol{A}^T\boldsymbol{A\mu}$$

- Here we use the Equation 3, we can simply get the result that:

$$Cov[g(\boldsymbol{x})] = Cov[\boldsymbol{Ax}, \boldsymbol{Ax}] = \boldsymbol{A}Cov[\boldsymbol{x}, \boldsymbol{x}]\boldsymbol{A}^T = \boldsymbol{A}Var(\boldsymbol{x})\boldsymbol{A}^T = \boldsymbol{A}\sum\boldsymbol{A}^T$$

References