

5004 Homework 2

RONG Shuo

November 7, 2024

Question 1:

1. For each of the following functions $f(x_1, x_2)$, find all critical points (i.e, all x_1, x_2 such that $\nabla f(x_1, x_2) = \mathbf{0}$).

- (a) $f(x_1, x_2) = (4x_1^2 - x_2)^2$
- (b) $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2x_2$
- (c) $f(x_1, x_2) = (x_1 - 2x_2)^4 + 64x_1x_2$
- (d) $f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 - x_2$

Answer :

(a)

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 2(4x_1^2 - x_2)(8x_1) = 16x_1(4x_1^2 - x_2) \\ \frac{\partial f}{\partial x_2} &= 2(4x_1^2 - x_2)(-1) = -2(4x_1^2 - x_2)\end{aligned}$$

set the gradient to 0:

$$16x_1(4x_1^2 - x_2) = 0 \tag{1}$$

$$-2(4x_1^2 - x_2) = 0 \tag{2}$$

if $x_1 = 0$, from (2) we can get that $x_2 = 0$ if $x_1 \neq 0$, from equation (1), we know:

$$4x_1^2 = x_2$$

this satisfied $(x_1, x_2) = (0, 0)$ Thus, we can conclude that the critical points are:

$$(x_1, x_2) = (x_1, 4x_1^2) \forall x_1 \in \mathbb{R}.$$

(b)

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 6x_1x_2 \\ \frac{\partial f}{\partial x_2} &= 6x_2^2 - 12x_2 + 3x_1^2\end{aligned}$$

set the gradient to 0:

$$\begin{aligned} 6x_1x_2 &= 0 \\ 6x_2^2 - 12x_2 + 3x_1^2 &= 0 \end{aligned}$$

if $x_1 = 0$,

$$6x_2^2 - 12x_2 = 0 \Rightarrow 6x_2(x_2 - 2) = 0$$

we can conclude that $(x_1, x_2) = (0, 0)$, or $(x_1, x_2) = (0, 2)$. if $x_2 = 0$,

$$3x_1^2 = 0 \Rightarrow x_1 = 0$$

This gives $(x_1, x_2) = (0, 0)$

In conclusion, the critical points is $(0, 0)$ and $(0, 2)$

(c)

$$\begin{aligned} \frac{f}{\partial x_1} &= 4(x_1 - 2x_2)^3 + 64x_2 \\ \frac{f}{\partial x_2} &= -8(x_1 - 2x_2)^3 + 64x_1 \end{aligned}$$

set the gradient to 0:

$$\begin{aligned} 4(x_1 - 2x_2)^3 + 64x_2 &= 0 \\ (x_1 - 2x_2)^2 &= -16x_2 \\ -8(x_1 - 2x_2)^3 + 64x_1 &= 0 \\ (x_1 - 2x_2)^3 &= 8x_1 \end{aligned}$$

$$\begin{aligned} (x_1 - 2x_2)^2 &= -16x_2 \\ (x_1 - 2x_2)^3 &= 8x_1 \end{aligned}$$

Question 2:

2. Find the gradient of the following functions, where the space \mathbb{R} and $\mathbb{R}^{n \times n}$ are equipped with the standard inner product.

(a) $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $(\lambda > 0)$ are given.

(b) $f(\mathbf{X}) = \mathbf{b}^T \mathbf{X} \mathbf{c}$, where $\mathbf{X} \in \mathbb{R}^{n \times n}$ and $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$

(c) $f(\mathbf{X}) = \mathbf{b} \mathbf{X}^T \mathbf{X} \mathbf{c}$, where $\mathbf{X} \in \mathbb{R}^{n \times n}$ and $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$

Answer :

Question 3:

3. Let $\{\mathbf{x}_i, y_i\}_{i=1}^N$ be given with $\mathbf{x}_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$. Assume $N < n$. Consider the ridge regression

$$\text{minimize}_{\mathbf{a} \in \mathbb{R}^n} \sum_{i=1}^N (\langle \mathbf{a}, \mathbf{x}_i \rangle - y_i)^2 + \lambda \|\mathbf{a}\|_2^2,$$

where $\lambda \in \mathbb{R}$ is a regularization parameter, and we set the bias $b = 0$ for simplicity.

(a) Prove that the solution must be in the form of $\mathbf{a} = \sum_{i=1}^N c_i \mathbf{x}_i$ for some $\mathbf{c} = [c_1, c_2, \dots, c_N]^T \in \mathbb{R}^N$.

(hint: similar to the proof of the representer theorem.)

(b) Re-express the minimization in terms of $c \in \mathbb{R}^N$, which has fewer unknowns than the original formulation as $N < n$.

Answer :

Question 4:

4. Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric positive semidefinite matrix, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

(a) Prove that \mathbf{x} is a global minimizer of f if and only if $\mathbf{A} \mathbf{x} = -\mathbf{b}$.

(b) Prove that f is bounded below over \mathbb{R}^n if and only if $\mathbf{b} \in \{\mathbf{A} \mathbf{y} : \mathbf{y} \in \mathbb{R}^n\}$.

Answer :

Question 5:

5. We consider the following optimization problem:

$$\text{minimize}_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = \log \left(\sum_{i=1}^m \exp(\mathbf{a}_i^T \mathbf{x} + b_i) \right) \quad (3)$$

where $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, \dots, b_m \in \mathbb{R}$ are given.

(a) Find the gradient of $f(\mathbf{x})$.

(b) If we use gradient descent to solve Problem (1), will it converge to the global minimizer? Please justify your answer.

Answer