Problem Set 5

Problems 1-4 correspond to "Unconstrained optimization I"

Problem 1

1/1 point (graded)

Let F be a function from $\mathbb{R}^d$ to $\mathbb{R}$. Which of the following is the most accurate description of the derivative ∇F?

○ It is a real number.

○ It is a d-dimensional vector.

○ For any point $u \in \mathbb{R}^d$, the derivative at that point, $\nabla F(u)$, is a real number.

● For any point $u \in \mathbb{R}^d$, the derivative at that point, $\nabla F(u)$, is a d-dimensional vector.

✓

Submit

Answers are displayed within the problem

Problem 2

6/6 points (graded)

Consider the following loss function on vectors $w \in \mathbb{R}^3$:

$$L(w) = w_1^2 - 2w_1w_2 + w_2^2 + 2w_3^2 + 3.$$

a) Compute $\nabla L(w)$. Match each of its coordinates to the following list:

Option 1: $4w_3$

Option 2: $2w_1 - 2w_2$

Option 3: $-2w_1 + 2w_2$

What is dL/dw_1? (Just answer 1,2,or 3)

2 ✓ **Answer:** 2

2

$dL/dw_2 =$

3 ✓ **Answer:** 3

3

$dL/dw_3 =$

1 ✓ **Answer:** 1

1

b) What is the minimum value of $L(w)$?

3 ✓ **Answer:** 3

3

c) Is there is a unique solution w at which this minimum is realized?

no ▼ ✓ **Answer:** no

d) Suppose we use gradient descent to minimize this function, and that the current estimate is $w = (1, 2, 3)$. If the step size is $\eta = 0.5$, what is the next estimate?

○ $w = (1, 1, 0)$

○ $w = (-1, 0, 1)$

● $w = (2, 1, -3)$

○ $w = (0, -1, -1.5)$

✓

Explanation

The derivative of $L(w)$ is $\nabla L(w) = (2w_1 - 2w_2, -2w_1 + 2w_2, 4w_3)$.
To minimize $L(w)$, we set the derivative to zero and get $w_1 = w_2$ and $w_3 = 0$. Thus there isn't a single minimizer, but rather infinitely many of them. The minimum value of $L(w)$, obtained at any of these points, is 3.
For the final part, let the current point be $w' = (1, 2, 3)$. The gradient *at this point* is $\nabla L(w') = (-2, 2, 12)$. Thus the gradient step updates w' to $w' - \eta \nabla L(w') = (1, 2, 3) - 0.5(-2, 2, 12) = (2, 1, -3)$.

Submit

Answers are displayed within the problem

Problem 3

1/1 point (graded)

We are given a set of data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, and we want to find a single point $z \in \mathbb{R}^d$ that minimizes the loss function

$$L(z) = \sum_{i=1}^{n} \|x^{(i)} - z\|^2.$$

Use calculus to determine z, in terms of the $x^{(i)}$. ($Hint:$ It might help to just start by looking at one particular coordinate.) Then select which of the following correctly describes the solution.

○ The sum of the $x^{(i)}$ vectors

● The average of the $x^{(i)}$ vectors

○ The average of the $x^{(i)}$ vectors, times a constant $c \neq 1$

○ Zero, regardless of what the $x^{(i)}$ vectors are

✓

Explanation

Notice that

$L(z) = \sum_{i=1}^{n} \left\| x^{(i)} - z \right\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{d} \left(x_j^{(i)} - z_j \right)^2.$

Take the derivative with respect to a single coordinate z_j:

$\frac{dL}{dz_j} = -2 \sum_{i=1}^{n} \left(x_j^{(i)} - z_j \right).$

Stacking these together into a single d-dimensional vector, we get

$\nabla L(z) = -2 \sum_{i=1}^{n} \left(x^{(i)} - z \right).$

Setting this to zero then yields the solution

$z = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}.$

Submit

Answers are displayed within the problem

Problem 4

2/2 points (graded)

Given a set of data points $x^{(1)}, \ldots, x^{(n)} \in \mathbb{R}^d$, we want to find the vector $w \in \mathbb{R}^d$ that minimizes this loss function:

$$L(w) = \sum_{i=1}^{n} \left(w \cdot x^{(i)} \right) + \frac{1}{2} c \left\| w \right\|^2.$$

Here $c > 0$ is some constant.

a) Let s denote the sum of the data points, that is, $s = \sum_{i=1}^{n} x^{(i)}$. Express $\nabla L(w)$ in terms of s, c, and w.

○ $\nabla L(w) = s + w$

● $\nabla L(w) = s + cw$

○ $\nabla L(w) = cw$

○ $\nabla L(w) = s/c + w$

✓

Answer
Correct: The derivative is $\nabla L(w) = \sum_i x^{(i)} + cw = s + cw$

b) What value of w minimizes $L(w)$? Give the answer in terms of s and c.

● $w = -\frac{s}{c}$

○ $w = cs$

○ $w = \frac{s}{4c}$

○ $w = -\frac{s}{2c}$

✓

Answer
Correct: This results from setting $\nabla L(w) = 0$.

Submit

Answers are displayed within the problem

Problems 5-7 correspond to "Convexity I"

Problem 5

7/7 points (graded)

For each of the following functions of one variable, say whether it is `convex`, `concave`, `both`, or `neither`.

a) $f\left(x\right)=x^2$

convex ▼ ✓ **Answer:** convex

Answer

Correct: $f''\left(x\right)=2$

b) $f\left(x\right)=-x^2$

concave ▼ ✓ **Answer:** concave

Answer

Correct: $f''\left(x\right)=-2$

c) $f\left(x\right)=x^2-2x+1$

convex ▼ ✓ **Answer:** convex

Answer

Correct: $f''\left(x\right)=2$

d) $f\left(x\right)=x$

both ▼ ✓ **Answer:** both

Answer

Correct: $f''\left(x\right)=0$

e) $f\left(x\right)=x^3$

neither ▼ ✓ **Answer:** neither

Answer

Correct: $f''\left(x\right)=6x$, which is sometimes positive, sometimes negative.

f) $f\left(x\right) = x^4$

convex ▼ ✓ **Answer:** convex

Answer

Correct: $f''\left(x\right) = 12x^2$

g) $f\left(x\right) = \ln x$

concave ▼ ✓ **Answer:** concave

Answer

Correct: $f''\left(x\right) = -1/x^2$

? **Hint (1 of 2):** First rule: a twice-differentiable function is convex if its second derivative is always ≥ 0.

Hint (2 of 2): Second rule: a function f is concave if and only if $-f$ is convex.

Next Hint

Submit

ℹ Answers are displayed within the problem

Problem 6

1/1 point (graded)

Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f\left(x_1, x_2, x_3\right) = x_1^2 + x_2^2 - x_3^2 - 4x_1x_2 + 6x_2x_3.$$

Compute and select the matrix of second derivatives (the Hessian) $H\left(x\right)$.

○ $\begin{pmatrix} 1 & -2 & 0 \\ -2 & 1 & 3 \\ 0 & 3 & -1 \end{pmatrix}$

○ $\begin{pmatrix} 2 & -4 & 0 \\ -4 & 2 & 6 \\ 0 & 6 & 0 \end{pmatrix}$

● $\begin{pmatrix} 2 & -4 & 0 \\ -4 & 2 & 6 \\ 0 & 6 & -2 \end{pmatrix}$

○ $\begin{pmatrix} 1 & -4 & 0 \\ 0 & 2 & 6 \\ 0 & 0 & -2 \end{pmatrix}$

✓

? Hint (1 of 1): Helpful first step: $\nabla f(x) = (2x_1 - 4x_2, 2x_2 - 4x_1 + 6x_3, -2x_3 + 6x_2)$ Next Hint

Submit

Answers are displayed within the problem

Problem 7

1/1 point (graded)

For some fixed vector $u \in \mathbb{R}^d$, define the function $F : \mathbb{R}^d \to \mathbb{R}$ by

$$F(x) = e^{u \cdot x}.$$

Which of the following is the Hessian $H(x)$?

● $e^{(u \cdot x)} uu^T$

○ $e^{(u \cdot x)} I$ (here I is the $d \times d$ identity matrix)

○ $e^{(u\cdot x)}\|u\|^2$

○ $e^{(u\cdot x)}(u\cdot x)^2$

✓

Explanation

First derivative:

$$\frac{dF}{dx_j} = e^{u\cdot x}u_j$$

Second derivative:

$$\frac{d^2F}{dx_k\,dx_j} = e^{u\cdot x}u_j u_k$$

Putting together the full $d \times d$ matrix, we get $e^{u\cdot x}uu^T$.

? **Hint (1 of 1):** Helpful first step: $\nabla F(x) = e^{u\cdot x}u$ Next Hint

Submit

ℹ Answers are displayed within the problem

Problems 8-11 correspond to "Positive semidefinite matrices"

Problem 8

1/1 point (graded)

Is the matrix $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ positive semidefinite?

○ Yes, because every entry in the matrix is ≥ 0

○ No, because not every entry is > 0

○ Yes, because $u^T Mu \geq 0$ for all vectors u

● No, because there is a vector u for which $u^T Mu < 0$

✓

Explanation

The quadratic function represented by this matrix is $u^T Mu = 2u_1 u_2$. This is negative whenever $u_1 u_2$ is negative, for instance with $u = (1, -1)$.

Submit

ℹ Answers are displayed within the problem

Problem 9

1/1 point (graded)

Is the matrix $M = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ positive semidefinite?

○ No, because not every entry is ≥ 0

● Yes, because $u^T Mu \geq 0$ for all vectors u

○ No, because there is a vector u for which $u^T Mu < 0$

○ No, because there is a vector u for which $u^T Mu = 0$

✓

Explanation

The quadratic function represented by this matrix is $u^T Mu = u_1^2 - 2u_1 u_2 + u_2^2 = (u_1 - u_2)^2$. This is never negative.

Submit

🛈 Answers are displayed within the problem

Problem 10

1/1 point (graded)

For a fixed set of vectors $v^{(1)}, \dots, v^{(n)} \in \mathbb{R}^d$, let M be the $n \times n$ matrix of all pairwise dot products: that is, $M_{ij} = v^{(i)} \cdot v^{(j)}$. Do you see why M is positive semidefinite? Think about it a little bit, and then choose one of the following options (you'll get marked as correct whichever you choose).

○ Yes, the entire argument is clear to me. ✓

○ That sounds right, but I can't fully construct the argument. ✓

● I don't get it.

✓

Explanation

Let U denote the $n \times d$ matrix whose rows are the $v^{(i)}$. Then $M = UU^T$, and thus M is PSD (any matrix that can be written in this way is PSD).

Submit

🛈 Answers are displayed within the problem

Problem 11

1/1 point (graded)

Suppose M and N are positive semidefinite matrices of the same size. Which of the following matrices are *necessarily* positive semidefinite? Select all that apply.

☑ $M + N$

☐ $M - N$

☑ $2M$

☑ $(1/2)\,M$

☑ M^TNM

✓

Explanation

The first is PSD because the sum of PSD matrices is also PSD. The third and fourth are PSD because any non-negative multiple of a PSD matrix is PSD.

The second option is *not* PSD: consider, for instance, the 1×1 matrices $M = 1$ and $N = 10$.

For the fourth option, notice that since N is PSD, we can write it in the form $N = UU^T$ for some matrix U. Then,

$$M^TNM = M^TUU^TM = (M^TU)\,(M^TU)^T = VV^T,$$

where $V = M^TU$. Thus this matrix is also PSD.

Submit

Answers are displayed within the problem

Problems 12-13 correspond to "Convexity II"

Problem 12

2/2 points (graded)

For some fixed vector $u \in \mathbb{R}^d$, define

$$F(x) = \|x - u\|^2.$$

We wish to determine whether $F(x)$ is a convex function of x.

a) The Hessian matrix $H(x)$ is of the form cI, where I is the $d \times d$ identity matrix and c is some constant. What is c?

2

✓ **Answer:** 2

2

b) Is $F(x)$ a convex function?

● Yes

○ No

○ It depends on the specific vector u

✓

Explanation
For the first part, we have
$F(x) = \sum_{j=1}^{d} (x_j - u_j)^2$.
Thus
$\frac{dF}{dx_j} = 2(x_j - u_j)$
and $d^2F/dx_k dx_j$ is either 2 if $j = k$ or 0 otherwise. Thus the Hessian is $2I$, which is PSD, implying that F is convex.

Submit

ℹ Answers are displayed within the problem

Problem 13

3/3 points (graded)
Let $p = (p_1, p_2, \ldots, p_m)$ be a probability distribution over m possible outcomes. The *entropy* of p is a measure of how much randomness there is in the outcome. It is defined as

$$F(p) = -\sum_{i=1}^{m} p_i \ln p_i,$$

where $\ln$ denotes natural logarithm. We wish to ascertain whether $F(p)$ is a convex function of p. As usual, we begin by computing the Hessian.

a) Consider the specific point $p = (1/m, 1/m, \ldots, 1/m)$. What is the $(1,1)$ entry of the Hessian at this point? Your answer should be a function of m.

-m

✓ **Answer:** -m

b) Continuing, what is the $(1,2)$ entry of the Hessian at this specific point?

0

✓ **Answer:** 0

0

c) Is the function $F(p)$ `convex`, `concave`, `both`, or `neither`?

concave ▼

✓ **Answer:** concave

Explanation

First we have

$$\frac{dF}{dp_i} = -(1 + \ln p_i)$$

Thus, if $j \neq i$, then

$$\frac{d^2F}{dp_j dp_i} = 0$$

while

$$\frac{d^2F}{dp_i^2} = -\frac{1}{p_i}.$$

Thus the Hessian is a diagonal matrix with negative entries, meaning the function is concave.

Submit

ℹ Answers are displayed within the problem