

Solution of the home assignment № 4

Andrei Ilyin

02.12.2024

1. Problem 1

What claims from below list are correct and what are incorrect and why?

1. Any convex function is smooth.
2. Any strongly convex function has a unique global minimum.
3. If a convex function is bounded below, then it has a unique point of minimum x^* .
4. A strongly convex function is always differentiable.

1) Any convex function is smooth

Incorrect. A convex function is not necessarily smooth. Convexity only requires that for any two points, the line segment joining them lies above the function graph, as defined by the inequality

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{for } \lambda \in [0, 1].$$

However, convex functions may be non-differentiable. For example, the absolute value function $f(x) = |x|$ is convex but not differentiable at $x = 0$.

2) Any strongly convex function has a unique global minimum

Correct. A strongly convex function satisfies the inequality

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}|y - x|^2,$$

where $m > 0$ is the strong convexity parameter. This property ensures strict convexity, which implies the global minimizer is unique if it exists.

3) If a convex function is bounded below, then it has a unique point of minimum x^*

Incorrect. A convex function can be bounded below without having a unique minimum. For example, the function $f(x) = 0$ for all x is convex, bounded below, and has infinitely many minima. Uniqueness of the minimum occurs only if the convex function is strictly convex or some additional criteria are met.

4) A strongly convex function is always differentiable

Incorrect. Strong convexity implies the function has certain curvature (i.e., it grows quadratically away from the minimum), but it does not necessarily imply differentiability. For example, a piecewise quadratic function can be strongly convex but non-differentiable at certain points (e.g., corners). Differentiability requires the gradient to exist everywhere.

2. Problem 2

What functions below are convex or concave and why?

1. $f(x) = \sup_{y \in C} h(y, x_i)$, where C is some given set. 1. $f(x) = \sup_{y \in C} h(y, x)$, where C is some given set.

Analysis: If $h(y, x)$ is convex in x for every $y \in C$, then $f(x)$, being the pointwise supremum of convex functions, is also convex. fmin.xyz

Conclusion: $f(x)$ is convex if $h(y, x)$ is convex in x for every $y \in C$.

2. $f(x) = \|Ax - b\|$, where $\|\cdot\|$ is an arbitrary norm.

Analysis: Norms are convex functions. Since $Ax - b$ is an affine transformation of x , the composition of a norm with an affine transformation is convex.

Conclusion: $f(x)$ is convex.

3. $f(x) = \min_{i=1, \dots, n} x_i$.

Analysis: The minimum of a set of linear functions $\{x_i\}$ is concave. This is because for any convex combination $\lambda x + (1 - \lambda)y$, we have:

$$\min_i (\lambda x_i + (1 - \lambda)y_i) \geq \lambda \min_i x_i + (1 - \lambda) \min_i y_i.$$

Conclusion: $f(x)$ is concave.

4. $f(x) = -(\prod_{i=1}^n x_i)^{\frac{1}{n}}$, $\text{dom } f = \mathbb{R}_+^n$.

Analysis: The function $(\prod_{i=1}^n x_i)^{1/n}$ is the geometric mean, which is concave over \mathbb{R}_+^n . Negating a concave function makes it convex.

Conclusion: $f(x)$ is convex.

5. $f(w) = \sum_{i=1}^m \log(1 + e^{-y_i w^\top x_i}) + \frac{1}{2} \|w\|_2^2$, where $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}$. This function is a basic loss for the binary classification problem.

Analysis: The logistic loss $\log(1 + e^{-y_i w^\top x_i})$ is convex in w because the second derivative is nonnegative. The ℓ_2 -regularization term $\frac{1}{2} \|w\|_2^2$ is also convex. The sum of convex functions is convex.

Conclusion: $f(w)$ is convex.

6. $f(X, Y) = \|A - XY\|_F^2$, where $X \in \mathbb{R}^{m \times p}$, $Y \in \mathbb{R}^{p \times n}$. The notation $\|\cdot\|_F$ means Frobenius norm, computed as follows: $\|X\|_F^2 = \sum_{i,j} x_{ij}^2$. The function f is the key ingredient of the loss in the matrix factorization model used in recommender systems. The matrix A is binary and represents the history of user-item interactions. Note that the convexity of f means that it is convex with respect to both arguments simultaneously.

Analysis: $f(X, Y)$ is quadratic in both (X) and (Y) but not jointly convex. Convexity with respect to both arguments simultaneously would require the Hessian to be positive semidefinite for all directions, which is generally not true for bilinear terms like (XY) .

Conclusion: $f(X, Y)$ is not convex with respect to (X) and (Y) simultaneously.

7. $f(W_1, W_2) = \|W_1 \max(W_2 x, 0)\|_2$, where \max is an element-wise function here. The vector x is given. The function inside the 2-norm is a toy instance of the DeepReLU neural network. Note that the convexity of f means that it is convex with respect to W_1 and W_2 simultaneously.

Analysis: $\max(W_2 x, 0)$ introduces nonlinearity that is not convex in general, and the composition with W_1 makes the analysis more complex. Even though the ℓ_2 -norm is convex, the product $W_1 \max(W_2 x, 0)$ is not guaranteed to be convex in W_1 and W_2 simultaneously.

Conclusion: $f(W_1, W_2)$ is not convex.

3. Problem 3

What claims from the list below are correct, what are incorrect and why?

1. Lipschitz constant of gradient bounds from above the norm of Hessian.
2. Lipschitz constant of gradient bounds from above the absolute values of function.
3. Lipschitz constant of function bounds from above the norm of Hessian.
4. Lipschitz constant of function bounds from above the norm of gradient.

1. Lipschitz constant of gradient bounds from above the norm of the Hessian

Analysis: - If the gradient $\nabla f(x)$ is Lipschitz continuous, it means there exists $L > 0$ such that:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y.$$

This condition implies that the Hessian $H(x) = \nabla^2 f(x)$ exists (almost everywhere for $f \in C^2$) and satisfies:

$$\|H(x)\| \leq L \quad \forall x,$$

where $\|H(x)\|$ is the operator norm of the Hessian. - Therefore, the Lipschitz constant L of the gradient directly bounds the norm of the Hessian.

Conclusion: Correct. The Lipschitz constant of the gradient bounds the norm of the Hessian.

2. Lipschitz constant of gradient bounds from above the absolute values of function

Analysis: - The Lipschitz continuity of the gradient governs the smoothness of $f(x)$, but it does not directly bound $f(x)$ itself. $f(x)$ can have arbitrarily large values depending on its definition, while the gradient's Lipschitz constant only restricts how fast $\nabla f(x)$ changes. -

Thus, there is no direct relationship between the Lipschitz constant of the gradient and the absolute values of $f(x)$.

Conclusion: Incorrect. The Lipschitz constant of the gradient does not bound the absolute values of $f(x)$.

3. Lipschitz constant of function bounds from above the norm of Hessian

Analysis: - The Lipschitz continuity of $f(x)$ implies that:

$$|f(x) - f(y)| \leq L\|x - y\| \quad \forall x, y,$$

where L is the Lipschitz constant of $f(x)$. However, this does not imply any relationship with the Hessian $\nabla^2 f(x)$, as the Hessian relates to second-order behavior, not first-order continuity. In particular, the Lipschitz constant of $f(x)$ does not provide any bound on $\|H(x)\|$.

Conclusion: Incorrect. The Lipschitz constant of $f(x)$ does not bound the norm of the Hessian.

4. Lipschitz constant of function bounds from above the norm of gradient

Analysis: - The Lipschitz constant L of $f(x)$ bounds the rate of change of $f(x)$, but it does not necessarily bound the norm of the gradient $\|\nabla f(x)\|$. - For example, even if $f(x)$ is Lipschitz continuous with constant L , its gradient norm $\|\nabla f(x)\|$ can exceed L . The bound is only on the *difference* of function values, not directly on the gradient norm.

Conclusion: Incorrect. The Lipschitz constant of $f(x)$ does not bound the norm of the gradient.