Machine Learning, 2021 Spring
Homework 2 and Solution

Due on 23:59 MAR 28, 2021

Problem 1

Prove that $f : \mathbb{R}^n \to \mathbb{R}$ is *affine* if and only if f is both convex and concave. [2pts]

Solution 1

$\Rightarrow$: If f is affine, then it has form $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\forall \theta \in [0,1]$

$$\begin{aligned} f\left(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) &= \boldsymbol{w}^T\left(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) + b & (1)\\ &= \theta\left(\boldsymbol{w}^T\boldsymbol{x} + b\right) + (1-\theta)\left(\boldsymbol{w}^T\boldsymbol{y} + b\right) & (2)\\ &= \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) & (3)\end{aligned}$$

the equality means both the convexity and concavity.
$\Leftarrow$: If f is both convex and concave, according to the definition, we have $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\forall \theta \in [0,1]$

$$f\left(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \le \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \le f\left(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \tag{4}$$

where the first inequality is from the convexity, the second inequality is from the concavity. Thus we have

$$f\left(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) = \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \tag{5}$$

let $\boldsymbol{y} = 0$, we have

$$f(\theta\boldsymbol{x}) = \theta f(\boldsymbol{x}) \tag{6}$$

let $\theta = \frac{1}{2}$, we have

$$f\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\right) = \frac{1}{2} f(\boldsymbol{x}+\boldsymbol{y}) = \frac{1}{2} f(\boldsymbol{x}) + \frac{1}{2} f(\boldsymbol{y}) \Rightarrow f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y}) \tag{7}$$

where the first equation is from (6). Function satisfies (6) and (7) is a linear function, which has the form : $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$, thus is affine.[a]

[a] Further details are omitted, see definition of linear function.

Problem 2

Suppose A and B are both convex sets, prove that $C = A \cap B$ is also convex. [1pts]

Solution 2

$C \subseteq A$ means that $\forall x, y \in C$, $x, y \in A$, similarly, $x, y \in B$. Thus, $\forall \theta \in [0, 1]$, by the convexity of A and B, we have:

$$\left.\begin{array}{l} \theta x + (1-\theta) y \in A \\ \theta x + (1-\theta) y \in B \end{array}\right\} \Rightarrow \theta x + (1-\theta) y \in A \cap B = C \tag{8}$$

which implies that C is also convex.

Problem 3

Suppose your algorithm for solving the problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) \tag{9}$$

takes iteration:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{p}^k \tag{10}$$

where $\boldsymbol{p}^k = \boldsymbol{H}^k \nabla f(\boldsymbol{x}^k)$. What kind of $\boldsymbol{H}^k$ can guarantee that $\boldsymbol{p}^k$ is a descent direction ? [2pts]

Solution 3

$\boldsymbol{p}^k$ is a descent direction if and only if $\langle \boldsymbol{p}^k, \nabla f(\boldsymbol{x}^k) \rangle < 0$, which is equivalent to

$$\langle \boldsymbol{H}^k \nabla f(\boldsymbol{x}^k), \nabla f(\boldsymbol{x}^k) \rangle = \nabla f(\boldsymbol{x}^k)^T \boldsymbol{H}^k \nabla f(\boldsymbol{x}^k) < 0 \tag{11}$$

thus when $\boldsymbol{H}^k$ is **negative definite**, $\boldsymbol{p}^k$ is a descent direction.

Problem 4

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. For a given $\boldsymbol{x} \in \mathbb{R}^n$, show that moving along $-\nabla f(\boldsymbol{x}) \neq 0$ with sufficiently small stepsize causes decrease on f, that is,

$$f(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})) < f(\boldsymbol{x}) \tag{12}$$

for sufficiently small $\alpha > 0$. [2pts]

Solution 4

By Taylor Theorem, we have

$$f(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})) = f(\boldsymbol{x}) - \alpha \|\nabla f(\boldsymbol{x})\|_2^2 + o(\alpha) \tag{13}$$

where $o(\alpha)$ is the residual term such that $\lim_{\alpha \to 0} \frac{o(\alpha)}{\alpha} = 0$. Thus for sufficiently small α,

$$o(\alpha) < \frac{1}{2} \alpha \|\nabla f(\boldsymbol{x})\|_2^2 \tag{14}$$

which means that

$$f(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})) < f(\boldsymbol{x}) \tag{15}$$

Problem 5

Use gradient descent to solve the *underdetermined* linear system:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{16}$$

with stepsize chosen as *exact line search*, initial point $\boldsymbol{x}^0 = 0$ and maximum iteration 1000. Plot :

1. The objective value against the iteration.(Use log scale for y-axis)
2. The ℓ_2 norm of gradient against the iteration.(Use log scale for y-axis)
3. The stepsize against the iteration.

The data $\boldsymbol{A} \in \mathbb{R}^{500\times 1000}$, $\boldsymbol{b} \in \mathbb{R}^{500\times 1}$ is attached in data/A.csv and data/b.csv with comma-separated (delimiter=','). [*Hint:* what is the solution to the exact line search for quadratic function?][3pts]

Solution 5

Let $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$, we first derive the solution of exact line search for least square problem:

$$\begin{aligned}
\alpha_k &= \arg\min_{\alpha > 0} f(\boldsymbol{x}^k - \alpha \nabla f(\boldsymbol{x}^k)) && (17)\\
&= \arg\min_{\alpha > 0} \left[\|\boldsymbol{A}\nabla f(\boldsymbol{x}^k)\|_2^2 \alpha^2 - 2\|\nabla f(\boldsymbol{x}^k)\|_2^2 \alpha + f(\boldsymbol{x}^k)\right] && (18)\\
&= \frac{\|\nabla f(\boldsymbol{x}^k)\|_2^2}{\|\boldsymbol{A}\nabla f(\boldsymbol{x}^k)\|_2^2} && (19)
\end{aligned}$$

The pesudo code for solving the LS problem is given in Algorithm [1]. The results are shown as follows:

Algorithm 1: Steepest Descent for Least Square problem

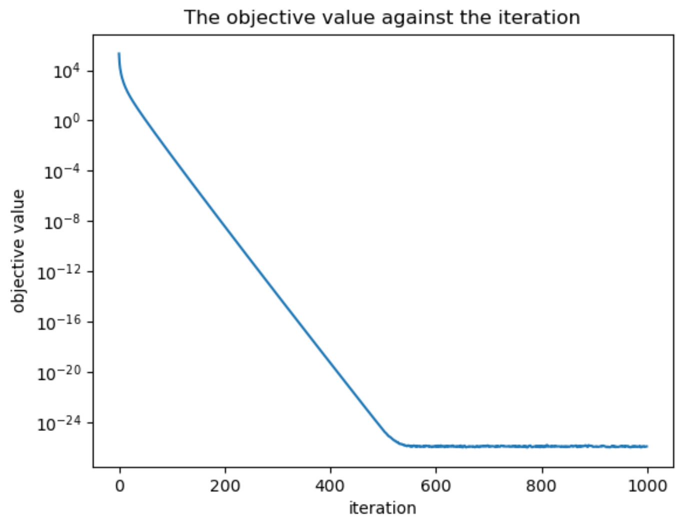
Input: $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, maximum iterations $M > 0$;
Initialization: $\boldsymbol{x}^0 = \boldsymbol{0}_{n\times 1}$;
for $k = 0, 2, \ldots, M-1$ **do**
  Gradient: $\nabla f(\boldsymbol{x}^k) = \boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b}\right)$;
  Exact line search:

$$\alpha_k = \frac{\|\nabla f(\boldsymbol{x}^k)\|_2^2}{\|\boldsymbol{A}\nabla f(\boldsymbol{x}^k)\|_2^2}$$

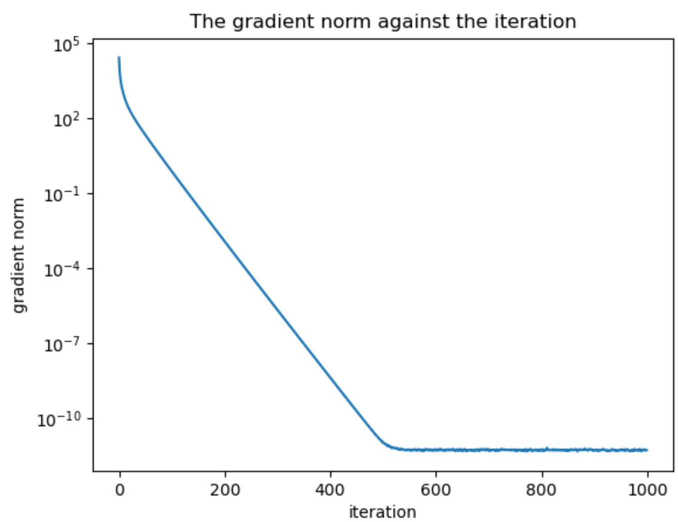
  Update the current point:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha \nabla f(\boldsymbol{x}^k)$$

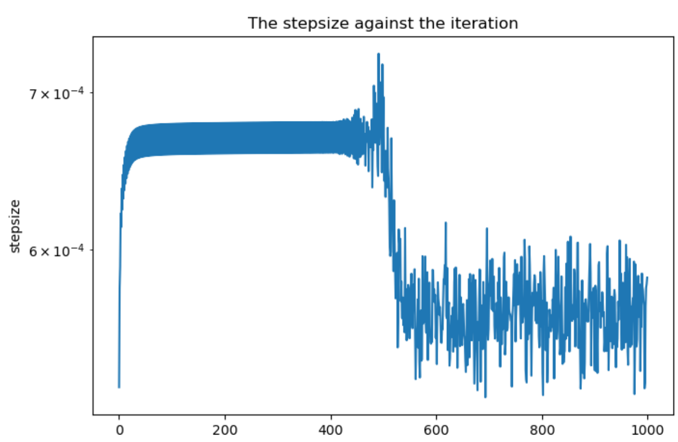
end
return $\boldsymbol{x}^M$.



The objective value against the iteration
objective value
iteration



The gradient norm against the iteration
gradient norm
iteration



The stepsize against the iteration
stepsize
iteration