**CSE 512: Homework** **Due Wednesday, October 6**

---

# Extra practice problems, ungraded

1. *Gradients.* Compute the gradients of the following functions. Give the exact dimension of the output.

   (a) *Linear regression.* $f(x) = \frac{1}{40}\|Ax-b\|_2^2$, $A \in \mathbb{R}^{20\times 10}$

   **Ans.** Actually, the best way to do this is to invoke the chain rule, which you will prove in the first graded problem. Write $g(v) = \frac{1}{40}\|v-b\|_2^2$. Then since $b \in \mathbb{R}^{20}$,

   $$\nabla g(v) = \nabla_v \left( \frac{1}{40} \sum_{i=1}^{20} (v_{[i]} - b_{[i]})^2 \right) \overset{\text{linearity}}{=} \frac{1}{40} \sum_{i=1}^{20} \nabla_v \left( (v_{[i]} - b_{[i]})^2 \right).$$

   Note that

   $$\nabla_v (v_{[i]} - b_{[i]})^2 = \begin{bmatrix} \frac{\partial}{\partial v_{[1]}} (v_{[i]} - b_{[i]})^2 \\ \frac{\partial}{\partial v_{[2]}} (v_{[i]} - b_{[i]})^2 \\ \vdots \\ \frac{\partial}{\partial v_{[20]}} (v_{[i]} - b_{[i]})^2 \end{bmatrix}$$

   and

   $$\frac{\partial}{\partial v_{[k]}} (v_{[i]} - b)^2 = \begin{cases} 2(v_{[i]} - b) & \text{if } i = k \\ 0 & \text{else.} \end{cases}$$

   So,

   $$\sum_{i=1}^{20} \nabla_v (v_{[i]} - b_{[i]})^2 = 2 \begin{bmatrix} \frac{\partial}{\partial v_{[1]}} (v_{[1]} - b_{[1]}) \\ \frac{\partial}{\partial v_{[2]}} (v_{[2]} - b_{[2]}) \\ \vdots \\ \frac{\partial}{\partial v_{[20]}} (v_{[20]} - b_{[20]}) \end{bmatrix} = 2(v-b).$$

   and $\nabla g(v) = \frac{1}{20}(v-b)$.
   Now, we invoke the chain rule. (Note that $f$ and $g$ are flipped as to their position in 1.(b).) Then

   $$\nabla f(x) = A^T \nabla g(Ax) = A^T(\frac{1}{20}(Ax-b)) = \frac{1}{20} A^T (Ax-b).$$

   To get the dimension, you can do this in two ways. One, you notice that $A$ has 10 columns, so $A^T$ has 10 rows. Two, you notice that the gradient $\nabla f(x)$ should always have the same number of elements as $x$, which is 10. In either case, $\nabla f(x) \in \mathbb{R}^{10}$.

   (b) *Sigmoid.* $f(x) = \sigma(c^T x)$, $c \in \mathbb{R}^5$, $\sigma(s) = \frac{1}{1+\exp(-x)}$. Hint: Start by showing that $\sigma'(s) = \sigma(s)(1-\sigma(s))$.

   **Ans.** We start with the hint, noting that

   $$\sigma'(s) = \frac{-\exp(-x)}{(1-\exp(-x))^2} = \frac{1}{1-\exp(-x)} \cdot \left(1 - \frac{1}{1-\exp(-x)}\right) = \sigma(s)(1-\sigma(s)).$$

   Then using chain rule, (where $A = c^T$) we can get

   $$\nabla f(x) = \sigma'(c^T x)c = \sigma(c^T x)(1 - \sigma(c^T x))c \in \mathbb{R}^5.$$

2. *Convex or not convex.* Are the following sets convex or not convex? Justify your answer.

(a) $\mathcal{S} = \textbf{range}(A) := \{x : Az = x \text{ for some } z\}$
**Ans.** This set is convex. Again, we check the definition: suppose $x \in \textbf{range}(A)$ and $y \in \textbf{range}(A)$. Then there exists some $u$ and $v$ where $Au = x$ and $Av = y$. Then for any $z = \theta x + (1-\theta)y$,

$$z = \theta Au + (1-\theta)Av = A\underbrace{(\theta u + (1-\theta)v)}_{w} = Aw \in \textbf{range}(A).$$

(b) $\mathcal{S} = \{x : x \le -1\} \cup \{x : x \ge 1\}$ (Read: either $x \le -1$ or $x \ge 1$.)
**Ans.** This set is not convex. We can just take

$$x = 1, \quad y = -1, \quad \theta = 1/2$$

and

$$\theta x + (1-\theta)y = 0 \notin \mathcal{S}.$$

3. *Am I positive semidefinite?* A symmetric matrix $X$ is *positive semidefinite* if for all $u$, $u^T X u \ge 0$. For each of the following, either prove that the matrix is positive semidefinite, or find a counterexample.

(a) $X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
**Ans.** The key observation here is that $X$ can be factorized, e.g.

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

Therefore, taking $c = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$,

$$u^T X u = (u^T c)^2 \ge 0 \quad \text{for all } u.$$

Therefore $X$ **is positive semidefinite.**

(b) $X = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$
**Ans.** Any matrix with negative diagonal elements is **not positive semidefinite**. To see this, pick $u = [1, 0, 0]^T$. Then

$$u^T X u = -1 < 0.$$

4. *Convex or not convex.* (**1pt, 0.125 points each.**) From lecture, we know that there are three ways of checking whether a function is convex or not.

- For any function, we can check if it satisfies the **definition of convexity**:

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y), \quad \forall x, y, \quad \forall 0 \le \theta \le 1.$$

- For any differentiable function, we can check the **first-order condition**

$$f(x) - f(y) \ge \nabla f(y)^T (x - y).$$

- For any twice-differentiable function, we can check the **second-order condition**

$$\nabla^2 f(x) \text{ is positive semidefinite, i.e. } u^T \nabla^2 f(x) u \ge 0 \quad \forall u.$$

Use any of these ways to determine whether or not each function is convex. (You only need to use one of these rules per function. Pick the one you think gets you to the answer the fastest!)

(a) $f(x) = \frac{1}{2}(x_{[1]}^2 - 2x_{[1]}x_{[2]} + x_{[2]}^2)$

**Ans.** This function is convex. We can use any of the three conditions to check.

- **By definition.** To make life easier, we first observe that this function can be rewritten as

$$f(x) = \frac{1}{2}(x_{[1]} - x_{[2]})^2.$$

Then

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \frac{1}{2}(\theta x_{[1]} + (1-\theta)y_{[1]} - \theta x_{[2]} - (1-\theta)y_{[2]})^2 \\ &= \frac{\theta^2}{2}(x_{[1]} - x_{[2]})^2 + \frac{(1-\theta)^2}{2}(y_{[1]} - y_{[2]})^2 - \theta(1-\theta)(x_{[1]} - x_{[2]})(y_{[1]} - y_{[2]}) \end{aligned}$$

Note that in general, since $(a+b)^2 \geq 0$, then $a^2 + b^2 \geq -2ab$, for any scalars $a$ and $b$. Therefore

$$(x_{[1]} - x_{[2]})(y_{[1]} - y_{[2]}) \leq \frac{1}{2}(x_{[1]} - x_{[2]})^2 + \frac{1}{2}(y_{[1]} - y_{[2]})^2.$$

Then

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leq \frac{\theta^2}{2}(x_{[1]} - x_{[2]})^2 + \frac{(1-\theta)^2}{2}(y_{[1]} - y_{[2]})^2 + \frac{1}{2}\theta(1-\theta)((x_{[1]} - x_{[2]})^2 + (y_{[1]} - y_{[2]})^2) \\ &= \frac{\theta^2 + \theta(1-\theta)}{2}(x_{[1]} - x_{[2]})^2 + \frac{(1-\theta)^2 + \theta(1-\theta)}{2}(y_{[1]} - y_{[2]})^2 \\ &= \frac{\theta}{2}(x_{[1]} - x_{[2]})^2 + \frac{1-\theta}{2}(y_{[1]} - y_{[2]})^2 \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

which satisfies the definition of a convex function.

- **By first order condition** Here, I'm going to again rewrite the problem as

$$f(x) = \frac{1}{2}(x^T c)^2$$

where $c = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then

$$\nabla f(x) = (x^T c)c$$

and

$$\begin{aligned} f(x) - f(y) - \nabla f(y)^T(x-y) &= \frac{1}{2}(x^T c)^2 - \frac{1}{2}(y^T c)^2 - (y^T c)c^T(x-y) \\ &= \frac{1}{2}(x^T c)^2 + \frac{1}{2}(y^T c)^2 - (y^T c)c^T x \\ &= \frac{1}{2}(x^T c - y^T c)^2 \geq 0, \quad \forall x, y. \end{aligned}$$

Thus the first order condition is satisfied!

- **By second order condition.** This is by far the fastest way to check.

$$\nabla^2 f(x) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

To show this matrix is positive semidefinite, it suffices to use the definition

$$u^T \nabla^2 f(x) u = (u_{[1]} - u_{[2]})^2 \geq 0, \quad \forall u.$$

Alternatively, we can try to "factorize" the Hessian, e.g. noticing that

$$\nabla^2 f(x) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T.$$

Therefore, taking $c = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, then

$$u^T \nabla^2 f(x) u = (c^T u)^2 \geq 0, \quad \forall u.$$

(b) $f(x) = |x|$ Hint: Again, remember the triangle inequality.

**Ans.** Since this function is not differentiable, our only option is to use the definition of convexity. We check as

$$f(\theta x + (1-\theta)y) = |\theta x + (1-\theta)y| \overset{\text{triangle inequality}}{\leq} |\theta x| + |(1-\theta)y| = \theta|x| + (1-\theta)|y| = \theta f(x) + (1-\theta) f(y).$$

Thus the definition of convex function is satisfied!

(c) $f(x) = \log(\exp(x_{[1]}) + \exp(x_{[2]}))$

**Ans.** This function is convex. For this problem, I really don't recommend trying to use the definition or first order condition, as they are really not straightforward to verify. However, the second order condition works nicely.

First, note we can write the gradient as

$$\nabla f(x) = \begin{bmatrix} \frac{\exp(x_{[1]})}{\exp(x_{[1]})+\exp(x_{[2]})} \\ \frac{\exp(x_{[2]})}{\exp(x_{[1]})+\exp(x_{[2]})} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\exp(x_{[2]})}{\exp(x_{[1]})+\exp(x_{[2]})} \\ 1 - \frac{\exp(x_{[1]})}{\exp(x_{[1]})+\exp(x_{[2]})} \end{bmatrix}$$

Now the Hessian can be written as

$$\nabla^2 f(x) = \frac{\exp(x_{[1]})\exp(x_{[2]})}{(\exp(x_{[1]}) + \exp(x_{[2]}))^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Since $\frac{\exp(x_{[1]})\exp(x_{[2]})}{(\exp(x_{[1]})+\exp(x_{[2]}))^2} > 0$ for all $x$, and the matrix

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T$$

is positive semidefinite, then the Hessian is positive semidefinite, and the function is convex.

(d) $f(x) = c^T x$

**Ans.** Since this function is linear, it is also convex. Any of the following arguments work:

- Using the definition of convexity,

$$f(\theta x + (1-\theta)y) = c^T(\theta x + (1-\theta)y) = \theta c^T x + (1-\theta)c^T y = \theta f(x) + (1-\theta)c^T y.$$

- Using first order condition,

$$f(x) - f(y) - \nabla f(y)^T (x - y) = c^T x - c^T y - c^T(x - y) = 0 \geq 0.$$

- Technically you could also say that the Hessian is $\nabla^2 f(x) = 0$ which is positive semidefinite, since for any $u$, $u^T \nabla^2 f(x) u = 0 \geq 0$.