Section 04: Solutions

1. Convexity

We've seen multiple equivalent definitions of what it means for a function to be convex. For today we'll be primarily using this one:

Definition 1 (convex functions). *A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** on a set A if for all $x, y \in A$ and $\lambda \in [0, 1]$:*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

(a) Which of the following functions are convex? (Hint: draw a picture)
(i) $|x|$ (ii) $\cos(x)$ (iii) x^2

Solution:

$|x|$ and x^2 are both convex. $\cos(x)$ is not convex since we can draw a line at two points (from say $\frac{\pi}{2}$ to $2\pi + \frac{\pi}{2}$) that is not above the function.

Proof that $|x|$ is convex:

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= |\lambda x + (1 - \lambda)y| \\ &\leq \lambda|x| + (1 - \lambda)|y| \end{aligned}$$

Proof that x^2 is convex:
We begin by examining the inequality

$$\begin{aligned} (\lambda x + (1 - \lambda)y)^2 &\leq \lambda x^2 + (1 - \lambda)y^2 \\ \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 &\leq \lambda x^2 + (1 - \lambda)y^2 \\ \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 - \lambda x^2 - (1 - \lambda)y^2 &\leq 0 \\ (\lambda^2 - \lambda)x^2 - 2(\lambda^2 - \lambda)xy + (\lambda^2 - \lambda)y^2 &\leq 0 \\ (\lambda^2 - \lambda)(x^2 - 2xy + y^2) &\leq 0 \\ (\lambda^2 - \lambda)(x - y)^2 &\leq 0 \end{aligned}$$

Which holds when $\lambda \in [0, 1]$, so the inequality is valid and our function is convex.

(b) Suppose you know that f and g are convex functions on a set A. Show that $h(x) := \max\{f(x), g(x)\}$ is also convex of A.

Solution:

$$\begin{aligned} h(\lambda x + (1 - \lambda)y) &= \max\{f(\lambda x + (1 - \lambda)y), g(\lambda x + (1 - \lambda)y\} && \text{Def of } h \\ &\leq \max\{\lambda f(x) + (1 - \lambda)f(y), \lambda g(x) + (1 - \lambda)g(y)\} && \text{Def of convexity} \\ &\leq \lambda \max\{f(x), g(x)\} + (1 - \lambda)\max\{f(y), g(y)\} && (*) \\ &= \lambda h(x) + (1 - \lambda)h(y) && \text{Def of } h \end{aligned}$$

$$(*) \; ||[a + b, c + d]||_\infty = ||[a, c] + [b, d]||_\infty \leq ||[a, c]||_\infty + ||[b, d]||_\infty$$

(c) Does the same result hold for $h(x) = \min\{f(x), g(x)\}$? If so, give a proof. If not, provide convex functions f, g such that h is not convex.

Solution:

No, consider $f(x) = x^2$, $g(x) = (x-1)^2$. Then $h(0) = h(1) = 0$, but $h(0.5) = 0.25$, so $h(0.5*0+0.5*1) = 0.25 > 0 = 0.5*h(0) + 0.5*h(1)$. So The minimum of two convex functions is not convex in general.

(d) Convex functions are useful because local minima are always global minima. Informally explain why this has to be true (a picture might help)

Solution:

Suppose we have a point x, which is a local minimum but not a global minimum. Since the function is convex, if we draw a line segment between x and a global minimum, the segment should be above the function. But the segment should be going down as it leaves the x (since the global minimum is overall lower than the function is at x) while f should be going up in every direction away from x (since x is a minimum) so the line segment is going down while f is going up, and the segment has to go below f, contradicting convexity.

2. Other Definitions of Convexity

Recall from the homework, that we also sometimes talk about a **set** being convex:

Definition 2 (convex set)**.** *A set A is **convex** if for all $x, y \in A$ and all $\lambda \in [0,1]$, the point $\lambda x + (1-\lambda)y$ is also in A.*

(a) On the homework, you use prove statements related to the following fact:

a function f defined on a convex set $A \subseteq \mathbb{R}^n$ is convex on A if and only if the set $S = \{(x, z) \in \mathbb{R}^{n+1} : z \geq f(x), x \in A\}$ is convex.

In this part, we'll prove the statement in general.

Solution:

Forward direction: suppose f is convex on a convex set A. We need to show $S = \{(x, z) : z \geq f(x), x \in A\}$ is convex.

Consider two points $(x,a),(y,b) \in S$. Applying the definition of convexity, we know for all $\lambda \in [0,1]$, $\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y)$ Now consider some point $z = (\lambda x + (1-\lambda)y, \lambda a + (1-\lambda)b)$ on the line segment between (x,a) and (y,b). Note that since A is convex $(\lambda x + (1-\lambda)y) \in A$. Thus to show z is in S, it suffices to show $\lambda a + (1-\lambda)b$ is above f:

$$\begin{aligned}\lambda a + (1-\lambda b) &\geq \lambda f(x) + (1-\lambda)f(y) \\ &\geq f(\lambda x + (1-\lambda)y)\end{aligned}$$

So $(\lambda x + (1-\lambda)y) \in S$, and S is convex.

Backward direction: Suppose $S = \{(x,z) : z \geq f(x), x \in A\}$ is convex. We need to show f is convex on A.
Let $x, y \in A$. By definition, $(x, f(x)), (y, f(y)) \in S$, so since S is convex, $(\lambda x+(1-\lambda)y, \lambda f(x)+(1-\lambda)f(y)) \in S$. By definition of S, $\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y)$, and f is convex.

(b) Suppose A and B are convex sets. Is $A \cap B$ convex? Is $A \cup B$ convex? Either prove or give a counter-example.

Solution:

$A \cap B$ is convex. Let $x, y \in A \cap B$. Consider any point $z = \lambda x + (1-\lambda)y$. Since $x, y \in A$, $z \in A$ and since $x, y \in B$, $z \in B$, so $z \in A \cap B$.

$A \cup B$ need not be convex – consider two completely disjoint convex sets. The union isn't even connected, so it certainly can't be a convex set.

(c)

Definition 3 (concave functions). *We say a function f is* ***concave*** *on a set A if for all $x, y \in A$ and all $\lambda \in [0,1]$*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$$

Show that if $f(x)$ is convex on A then $g(x) := -f(x)$ is concave on A.

Solution:

Since $f(x)$ is convex, we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Multiplying by -1 we get:

$$-f(\lambda x + (1-\lambda)y) \geq \lambda - f(x) + (1-\lambda) - f(y)$$

which is the same as

$$g(\lambda x + (1-\lambda)y) \geq \lambda g(x) + (1-\lambda)g(y)$$

so g is concave.

(d) Can a function be both convex and concave on the same set? If so, give an example. If not, describe why not.

Solution:

Linear functions (i.e. functions such that $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$ are both convex and concave.

(e) There is another definition of convex functions if we know the function is twice differentiable:

Definition 4 (convexity, second order condition)**.** *A twice-differentiable function is convex if* $f''(x) \geq 0$ *for all* x

Use the definition above to show $-\log(x)$ is convex.

Solution:

$f'(x) = \frac{-1}{x}$, $f''(x) = -\frac{-1}{x^2} = \frac{1}{x^2}$. For $x > 0$ (when $-\log(x)$ is defined), $f''(x)$ is always positive, so $-\log(x)$ is convex.

As a sanity check, it should be easy to see from a plot that $\log(x)$ is concave, so $-\log(x)$ is convex.

(f) Use the fact that $-\log(x)$ is convex to show the "arithmetic mean-geometric mean (AMGM) inequality": If $a, b \geq 0$ then $\sqrt{ab} \leq \frac{a+b}{2}$.

Solution:

Since log is monotonically increasing, it suffices to show $\frac{\log(ab)}{2} = \frac{\log(a)+\log(b)}{2} \leq \log\left(\frac{a+b}{2}\right)$. Consider the points a, b. Setting $\lambda = 1/2$ in the definition of $-\log()$ being convex, we have:

$$-\log\left(\frac{1}{2}a + \frac{1}{2}b\right) \leq -\frac{1}{2}\log(a) + -\frac{1}{2}\log(b) = \frac{-(\log(a) + \log(b)))}{2}$$

Multiplying by -1 (and thus flipping the inequality) gives the desired statement.

(g) Show that if f is convex, then $g(x) = f(ax + b)$ (where a, b are real numbers) is also convex.

Solution:

We need to show for any x, y and any $\lambda \in [0, 1]$, $g(\lambda x + (1 - \lambda)y \leq \lambda g(x) + (1 - \lambda)g(y)$.

$$\begin{aligned} g(\lambda x + (1 - \lambda)y) &= f(a(\lambda x + (1 - \lambda)y) + b) \\ &= f(\lambda(ax + b) + (1 - \lambda)(ay + b)) \\ &\leq \lambda f(ax + b) + (1 - \lambda)f(ay + b) \\ &= \lambda g(x) + (1 - \lambda)g(y) \end{aligned}$$

3. Gradient Descent

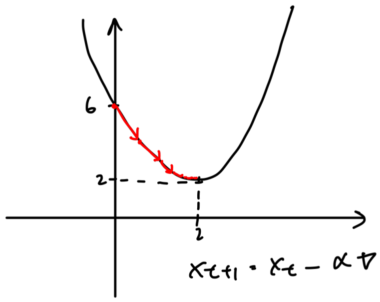
We will now examine gradient descent algorithm and study the effect of learning rate α on the convergence of the algorithm. Recall from lecture that Gradient Descent takes on the form of $x_{t+1} = x_t - \alpha \nabla f$

(a) Suppose we want to minimize the function $f(x) = x^2 - 4x + 6$. Run gradient descent by hand to compute using $\alpha = 0.5, 0.1, 10$. For each value of alpha, what was the observation? Don't worry about completing the computation, the goal here is for you to notice a trend following each α value picked. **Solution:**

For $\alpha = 0.5$, the algorithm reaches the minimum at 1 step. For $\alpha = 0.1$, the algorithm eventually converges, but takes more steps than $\alpha = 0.5$. For $\alpha = 10$, the algorithm diverges.

$\alpha = 0.1$

$x_1 = 6$



$x_{t+1} = x_t - \alpha \nabla y(x_t)$

$x_1 = 6$

$x_2 = 6 - 0.1\,(8) = 5.2$

$x_3 = 5.2 - 0.1\,(6.4) = 4.56$

$x_4 = 4.56 - 0.1\,(5.12) = 4.048$

$\vdots$

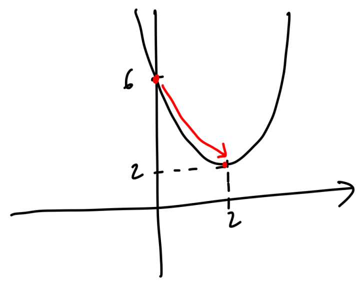
until you converges / any stopping condition

$y = x^2 - 4x + 6$

$\nabla y = 2x - 4$

$\alpha = 0.5$

$x_1 = 6$



$x_{t+1} = x_t - \alpha \nabla y(x)$

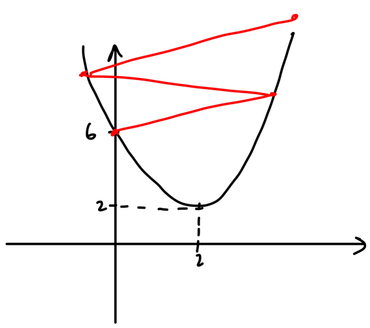
$x_1 = 6$

$x_2 = 6 - 0.5\,(8) = 2$

If converges in a single step, which is not always the case. In fact, it is rare that this happens.

$\alpha = 10$

$x_1 = 6$



Note that if you follow the computation, the algorithm will cause you to move further away from the minimum.

(b) Consider the two variable function $f(x, y) = x^2y^2 + x^2 - 10x$. Starting from the point $(2, 3)$ run gradient descent with a step size of 0.1.

Solution:

$\nabla(f) = \left[2xy^2 + 2x - 10, 2x^2y\right]^T$

$(x_1, y_1) = (2, 3)$
$(x_2, y_2) = (2, 3) - 0.1(30, 24) = (-1, 0.6)$
$(x_3, y_3) = (-1, 0.6) - 0.1(-12.72, 1.2) = (0.272, -0.6)$
$(x_4, y_4) \approx (0.272, -0.6) - 0.1(-9.26, -0.0888) \approx (1.20, -0.59)$

The process will continue from here, the minimum is $(5, 0)$.