Homework 6

by Mher Movsisyan

Convexity of functions

Problem 1:

Given n convex functions $g_i : R^{d_i} \to R \text{ for } i \in \{1, \dots, n\}$, prove or disprove that the function
a) $h(x) = g_2(g_1(x))$ is convex (here $d_1 \in N, d_2 = 1$)

Answer:

A function $f(x)$ is convex if for any $x, y \in R^{d_1}$ and any $\lambda \in [0, 1]$, we have:

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

So, for $h(x)$, we need to check whether:

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

Substitute the definition of $h(x)$:

$$g_2(g_1(\lambda x + (1-\lambda)y)) \leq \lambda g_2(g_1(x)) + (1-\lambda)g_2(g_1(y))$$

Now, let's use the fact that g_1 and g_2 are convex functions. Since g_1 is convex, we have:

$$g_1(\lambda x + (1-\lambda)y) \leq \lambda g_1(x) + (1-\lambda)g_1(y) \qquad (*)$$

And since g_2 is convex, for any $u, v \in R$ and any $\lambda \in [0, 1]$, we have:

$$g_2(\lambda u + (1-\lambda)v) \leq \lambda g_2(u) + (1-\lambda)g_2(v)$$

Now, let $u = g_1(x)$ and $v = g_1(y)$. Plugging these into the inequality for g_2:

$$g_2(\lambda g_1(x) + (1-\lambda)g_1(y)) \leq \lambda g_2(g_1(x)) + (1-\lambda)g_2(g_1(y))$$

Notice that the left-hand side of the inequality for g_2 is equal to $g_2(g_1(\lambda x + (1-\lambda)y))$. Now, combining this inequality with the inequality for g_1 in $(*)$, we have:

$$g_2(g_1(\lambda x + (1-\lambda)y)) \leq \lambda g_2(g_1(x)) + (1-\lambda)g_2(g_1(y))$$

Which is precisely the inequality we need to prove for the convexity of $h(x)$. Thus, $h(x) = g_2(g_1(x))$ is indeed convex.

b) $h(x) = g_2(g_1(x))$ is convex if g_2 is non-decreasing (here $d_1 \in N, d_2 = 1$)

Answer:

Since the previous one was a more general case of this problem, we can infer that $h(x)$ is convex. :)

c) $h(x) = max(g_1(x), \ldots, g_n(x))$ is convex (here all $d_i \in N$).

Answer:

Since we have the proof from part `a`, we only need to prove that the maximum function is convex to prove that h is convex.
For any number $k, i \in \{1, \ldots, n\}$, we have

$$\lambda x_k + (1-\lambda) y_k \leq \lambda \underset{i}{max} x_i + (1-\lambda) \underset{i}{max} y_i$$

if $\lambda \in [0, 1]$, so by definition the maximum function is convex. An easier way to prove this would be to think of the maximum function as an intersection of convex sets.

Optimization / Gradient descent

Problem 2:

You are given the following objective function

$$f(x_1, x_2) = 0.5x_1^2 + x_2^2 + 2x_1 + x_2 + cos(sin(\sqrt{\pi}))$$

a) Compute the minimizer x^* of f analytically.

Answer:

$$\frac{df}{dx_1} = x_1 + 2$$

$$\frac{df}{dx_1} = 2x_2 + 1$$

The roots are [-2, $-\frac{1}{2}$]

$$\frac{df}{dx_1^2} = 1$$

$$\frac{df}{dx_1 dx_2} = 0$$

$$\frac{df}{dx_2 dx_1} = 0$$

$$\frac{df}{dx_2^2} = 2$$

We clearly see that the Hessian is a positive definite matrix, therefore the (-2, $-\frac{1}{2}$) point is the global minimum.

b) Perform 2 steps of gradient descent on f starting from the point $x^{(0)} = (0, 0)$ with a constant learning rate $\tau = 1$.

Answer:

$$\nabla(0, 0) = (2, 1)$$

$$(x_1^{(1)}, x_2^{(1)}) = (0, 0) - \tau * (2, 1) = (-2, -1)$$

$$\nabla(-2, -1) = (0, -1)$$

$$(x_1^{(2)}, x_2^{(2)}) = (-2, -1) - \tau * (0, -1) = (-2, 0)$$

c) Will the gradient descent procedure from `Problem b)` ever converge to the true minimizer x^*? Why or why not? If the answer is no, how can we fix it?

Answer:
It will never converge because the learning rate is too big. We can decay it gradually untill it does.