Problem 1.

In the gradient descent algorithm, $\alpha > 0$ is the learning rate. If α is small enough, then the function value guarantees to decrease. In practice, we may anneal α, meaning that we start from a relatively large α, but decrease it gradually.

Show that α cannot be decreased too fast. If α is decreased too fast, even if it is strictly positive, the gradient descent algorithm may not converge to the optimum of a convex function.

Hint: Show a specific loss and an annealing scheduler such that the gradient descent algorithm fails to converge to the optimum.

Suppose we have a function $f(x) = wx$, and we want to optimize it using L_1 loss: $J(w) = |f(x) - t)|$. If we have only one data point in our dataset $\mathcal{D} = \{(1, 0)\}$, and we want to find the value w that fits this dataset. In this case, it is easy to see that the global optimal is $w^* = 0$.

With the gradient descent algorithm, each time step t we update w with $w^{(t)} = w^{(t-1)} - \nabla_w J(w) = w^{(t-1)} - \alpha^{(t-1)} \nabla_w |w^{(t-1)}|$, where $w \neq 0$.

Suppose our gradient descent starts with $w^{(0)} = 1$, and it has a small initial learning rate $\alpha^{(0)} = 0.1$. Therefore, the gradient descent is converging to $w^* = 0$ from the $w > 0$ side. Thus, we have $\nabla_w |w^{(t-1)}| = 1$.

Now, let us use an annealed learning rate $\alpha^{(t)} = \frac{1}{2^t} \alpha^{(0)}$. The annealed gradient descent computes $w^{(t)}$ as following

$$w^{(t)} = w^{(0)} - \alpha^{(0)} (\frac{1}{2^0} + \frac{1}{2^1} + \cdots + \frac{1}{2^{t-1}}) \tag{1}$$

We know that

$$\lim_{t \to \infty} (\frac{1}{2^0} + \frac{1}{2^1} + \cdots + \frac{1}{2^{t-1}}) = 2 \tag{2}$$

Therefore,

$$\lim_{t \to \infty} w^{(t)} = w^{(0)} - 2\alpha^{(0)} = 0.8 > 0 \tag{3}$$

This example shows that, a decayed learning rate may prevents the gradient descent algorithm from having enough energy for finding better local/global optimums.

Problem 2

Consider the Monty Hall game in a TV show. There are three closed doors, behind which are a car and two goats placed randomly.

a) You are asked to open a door by the host. Say, you would like to open Door 1. What is the probability of getting a car?
   1/3

b) The host knows where the car is but he/she does not tell you. Instead, the host will open another door with a goat.
   i) If the car is behind Door 2, the host can only open Door 3.
   ii) If the car is behind Door 1, the host can open either Door 2 or Door 3. He/she will do it with equal probability.

   Say, the host has opened Door 3. What is the probability of having the car behind Door 1 now? What is the probability of having the car behind Door 2 now?

   Following the hint, calculating

   $P(A = 1|B = 1) = \frac{P(A=1)P(B=1|A=1)}{P(A=1)P(B=1|A=1)+P(A=2)P(B=1|A=2)+P(A=3)P(B=1|A=3)} = \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1/3 \cdot 1 + 1/3 \cdot 0} = 1/3.$

c) If my goal is to get the car, should I change my first choice (i.e., open Door 1 or Door 2)?

   Yes. The probability of winning a car if sticks to one's choice, as calculated above, is $1/3$ while the probability of winning a car if switches, is $1 - 1/3 = 2/3$. One can also calculate the value of $P(A = 2|B = 1)$ following the similar calculations detailed in the hint:

   $P(A = 2|B = 1) = \frac{P(A=2)P(B=1|A=2)}{P(A=1)P(B=1|A=1)+P(A=2)P(B=1|A=2)+P(A=3)P(B=1|A=3)} = \frac{1/3 \cdot 1}{1/2 \cdot 1/3 + 1/3 \cdot 1} = 2/3.$

Hint: Suppose Door 1 is chosen in the first stage. Consider the random variables

- $A \in \{1, 2, 3\}$ is the door that has a car
- $B \in \{0, 1\}$ is whether the host opens Door 3 with a goat

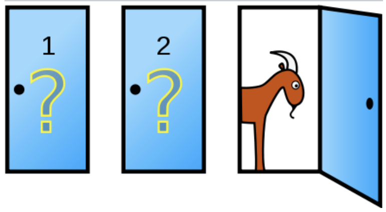
Based on Bayes' rule, we have

$$P(A = 1|B = 1) = \frac{P(A=1)P(B=1|A=1)}{P(A=1)P(B=1|A=1)+P(A=2)P(B=1|A=2)+P(A=3)P(B=1|A=3)}$$

where

- The numerator is essentially $P(A = 1, B = 1)$.
  - $P(A = 1) = 1/3$, the probability that the car is behind Door 1 before the game starts.
  - $P(B = 1|A = 1) = 1/2$, the probability that the host will open Door 3 if the car is behind Door 1.
- The denominator is essentially $P(B)$, the probability that the host opens Door 3 with a goat.
  - The first term $P(A = 1)P(B = 1|A = 1)$ is the same as numerator
  - For the second term $P(A = 2)P(B = 1|A = 2)$, we have
    - $P(A = 2) = 1/3$
    - $P(B = 1|A = 2) = 1$, because if the car is behind Door 2, the host can only open Door 3.
  - For the second term $P(A = 3)P(B = 1|A = 3)$, we have
    - $P(A = 3) = 1/3$
    - But $P(B = 1|A = 3) = 0$, because if the car is behind Door 3, there is no chance that the host can open the Door 3 with a goat, i.e., $B = 1$ cannot happen.

Now you can compute the probability of $P(A = 1|B = 1)$.



Reference:
https://en.wikipedia.org/wiki/Monty_Hall_problem#:~:text=Before%20the%20host%20opens%20a,%2F2%20%3D%201%2F6.

Problem 3.

Prove that the expectation is a linear system.

$$\mathbb{E}_{X\sim P(X)}[a\, f(X) \,+\, b\, g(X)] \,=\, a\, \mathbb{E}_{X\sim P(X)}[f(X)] \,+\, b\, \mathbb{E}_{X\sim P(X)}[g(X)]$$

$$\mathbb{E}_{X\sim P(X)}[a\, f(X) \,+\, b\, g(X)] \,=\, \sum_x P(x)[af(x) \,+\, bg(x)] \,=\, a\sum_x P(x)f(x) \,+\, b\sum_x P(x)g(x) \,=\, a\, \mathbb{E}_{X\sim P(X)}[f(X)] \,+\, b\, \mathbb{E}_{X\sim P(X)}[g(X)]$$

END