Assignment Zero - Solutions
ECE 4950

- This assignment will not be graded.
- You are expected to turn it in.
- Provide credit to **any sources** other than the course staff that helped you solve the problems.
- You can look up definitions/basics online (eg wikipedia, stackexchange, etc)
- Assignments will be uploaded on CMS.

Problem 1. A function f over a space $\mathcal{X}$ is convex if and only if for $0 \leq \lambda \leq 1$, and $x_1, x_2 \in \mathcal{X}$,

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2).$$

f is called mid-point convex if

$$\frac{f(x_1) + f(x_2)}{2} \geq f(\frac{x_1 + x_2}{2}).$$

All functions we will deal in this course are convex if and only if they are mid-point convex. Throughout this course, assume mid-point convexity to be equivalent to convexity.

A function f is concave if $-f$ is convex.

1. Let f be convex. Suppose $\lambda_1, \ldots, \lambda_k \geq 0$ such that $\lambda_1 + \ldots + \lambda_k = 1$. Prove that

$$\lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_k f(x_k) \geq f(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k).$$

Solution. We can prove it by induction. The case of $k = 2$ holds by the definition of convexity. Suppose $\lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_k f(x_k) \geq f(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k)$ holds for all $k \leq m$ (and $\lambda_1 + \ldots + \lambda_k = 1$). Now consider $k = m+1$, and $\lambda_1 + \ldots + \lambda_{m+1} = 1$. If $\lambda_{m+1} = 0$ this reduces to $k = m$. If $\lambda_{m+1} = 1$, all other λ_i's are 0, and we have equality. Hence assume that $0 < \lambda_{m+1} < 1$. Let $S_m = \lambda_1 + \ldots + \lambda_m$. Note that $\lambda_1 + \ldots + \lambda_{m+1} = S_m + \lambda_{m+1} = 1$. Therefore,

$$\begin{aligned}
&\lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_{m+1} f(x_{m+1}) && (1)\\
\geq\ & S_m \left(\frac{\lambda_1}{S_m} f(x_1) + \frac{\lambda_2}{S_m} f(x_2) + \ldots + \frac{\lambda_m}{S_m} f(x_m) \right) + \lambda_{m+1} f(x_{m+1}) \\
\geq\ & S_m \cdot f\left(\frac{\lambda_1}{S_m} x_1 + \frac{\lambda_2}{S_m} x_2 + \ldots + \frac{\lambda_m}{S_m} x_m \right) + \lambda_{m+1} f(x_{m+1}) && (2)\\
\geq\ & S_m \cdot f\left(\frac{\lambda_1}{S_m} x_1 + \frac{\lambda_2}{S_m} x_2 + \ldots + \frac{\lambda_m}{S_m} x_m) \right) + \lambda_{m+1} f(x_{m+1}) && (3)\\
\geq\ & f\left(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_m x_m + \lambda_{m+1} x_{m+1} \right), && (4)
\end{aligned}$$

where (2) is simply multiplying and dividing S_m, (3) follows from the induction hypothesis for $k = m$, and using $S_m = \lambda_1 + \ldots + \lambda_m$, (4) from the base case $k = 2$, and $S_m + \lambda_{m+1} = 1$.

2. If f and g are two convex functions, show that $\max\{f, g\}$, and $f + g$ are both convex.
Solution. The definition of $(f + g)(x)$ is $f(x) + g(x)$.

$$\begin{aligned}
\lambda \cdot (f+g)(x_1) + (1-\lambda) \cdot (f+g)(x_2) =& \lambda(f(x_1) + g(x_1)) + (1-\lambda)(f(x_2) + g(x_2)) && (5)\\
=& \lambda f(x_1) + (1-\lambda) f(x_2) + \lambda g(x_1) + (1-\lambda) g(x_2) && (6)\\
\geq& f(\lambda x_1 + (1-\lambda)x_2) + g(\lambda x_1 + (1-\lambda)x_2) && (7)\\
=& (f+g)(\lambda x_1 + (1-\lambda)x_2), && (8)
\end{aligned}$$

where the inequality is from convexity of f and g.

Suppose $h(x) = \max\{f(x), g(x)\}$. Then,

$$\begin{aligned}
\lambda h(x_1) + (1-\lambda) h(x_2) =& \lambda \max\{f(x_1), g(x_1)\} + (1-\lambda)\max\{f(x_2), g(x_2)\} && (9)\\
\geq& \lambda f(x_1) + (1-\lambda) f(x_2) && (10)\\
\geq& f(\lambda x_1 + (1-\lambda)x_2) && (11)
\end{aligned}$$

Similarly,

$$\lambda h(x_1) + (1-\lambda) h(x_2) \geq g(\lambda x_1 + (1-\lambda)x_2). \tag{12}$$

Therefore,

$$\begin{aligned}
\lambda h(x_1) + (1-\lambda) h(x_2) \geq& \max\{f(\lambda x_1 + (1-\lambda)x_2), g(\lambda x_1 + (1-\lambda)x_2)\} && (13)\\
=& h(\lambda x_1 + (1-\lambda)x_2) && (14)
\end{aligned}$$

3. Show that $\log(x)$ is concave over $(0, \infty)$.
Solution. Let $x, y > 0$. Then, $(\sqrt{x} - \sqrt{y})^2 \geq 0$, implying $\frac{x+y}{2} \geq \sqrt{xy}$. Since logarithm is monotonically increasing ($\log a > \log b$ if and only if $a > b$),

$$\log(\frac{x+y}{2}) \geq \log(\sqrt{xy}) = \frac{1}{2}(\log(x) + \log(y)).$$

Alternate Solution For functions that are differentiable, an alternate definition of convexity is $f''(x) \geq 0$ (and concavity is $f''(x) \leq 0$). Let $f(x) = \log x$, then $f''(x) = -1/x^2 < 0$.

4. Show that e^x and e^{-x} are both convex over $x \in \mathbb{R}$.
Solution. Since $(e^{x/2} - e^{y/2})^2 \geq 0$ for any x, y, we have

$$\frac{e^x + e^y}{2} \geq e^{(x+y)/2}. \tag{15}$$

Identical reasoning holds for e^{-x}.
Alternate Solution $f''(x) = f(x) > 0$ for $f(x) = e^x$, and $f(x) = e^{-x}$.

5. Let $x_1, x_2, y_1, y_2 > 0$. Then show that

$$x_1 \log \frac{x_1}{y_1} + x_2 \log \frac{x_2}{y_2} \geq (x_1 + x_2) \log \frac{x_1 + x_2}{y_1 + y_2}.$$

Solution. Since $\log(x)$ is concave, $-\log(x) = \log(1/x)$ is convex. Using these for $\lambda_1 = x_1/(x_1 + x_2)$, and $\lambda_2 = x_2/(x_1 + x_2)$,

$$\begin{aligned}
&\frac{x_1}{(x_1 + x_2)} \cdot \log \left(\frac{1}{y_1/x_1} \right) + \frac{x_2}{(x_1 + x_2)} \log \left(\frac{1}{y_2/x_2} \right) && (16) \\
\geq &\log \left(\frac{x_1}{(x_1 + x_2)} \cdot \frac{y_1}{x_1} + \frac{x_2}{(x_1 + x_2)} \cdot \frac{y_2}{x_2} \right) && (17) \\
= &\log \frac{x_1 + x_2}{y_1 + y_2}. && (18)
\end{aligned}$$

This is a rearrangement of the inequality in the problem.

Problem 2. A discrete distribution p over $\{1, \ldots, k\}$ is a map $p : \{1, \ldots, k\} \to [0, 1]$, such that $\sum_x p_x = 1$ (p_x is the probability of x under p). The entropy of p is $H(p) = \sum_x p_x \log \frac{1}{p_x}$.

1. Show that $0 \leq H(p) \leq \log k$. (assume $0 \log 0 = 0$)
   Solution. Recall that $\log x \geq 0$, for $x \geq 1$. Since $p_x \leq 1$, $1/p_x \geq 1$, implying that $p_x \log(1/p_x) \geq 0$. This shows that $H(p) \geq 0$.

   Using concavity of logarithms with $\lambda_i = p_i$, for $i = 1, \ldots, k$,

$$H(p) = \sum_{x=1}^{k} p_x \log \frac{1}{p_x} \leq \log \left(p_x \cdot \frac{1}{p_x} \right) = \log(k). \tag{19}$$

2. For any p, show that $H(p) \geq -\log(\sum_x p_x^2)$.
   Solution. By convexity of $-\log x$, and again using $\lambda_i = p_i$,

$$H(p) = \sum_{x=1}^{k} p_x \cdot (-\log p_x) \geq -\log(\sum_{x=1}^{k} p_x \cdot p_x).$$

3. For distributions p, q over $\{1, \ldots, k\}$, $(p + q)/2$ be the distribution that assigns probability $(p_x + q_x)/2$ to the symbol x. Show that $H((p + q)/2) \geq \frac{1}{2}(H(p) + H(q))$, namely entropy is concave.
   Solution. We prove the result term by term. Namely we will show:

$$\frac{p_x + q_x}{2} \log \frac{2}{p_x + q_x} \geq \frac{1}{2} \left(p_x \log \frac{1}{p_x} + q_x \log \frac{1}{q_x} \right).$$

   Summing over all x gives the result.

   In Problem 1, part 5 take $y_1 = y_2 = 1$, $x_1 = p_x$, and $x_2 = q_x$, and manipulate a bit to get the expression above.

Problem 3. The most ubiquitous continuous distributions are Gaussian distributions. They are parametrized by $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$, and are denoted $N(\mu, \sigma^2)$. Their probability density function is given by:

$$N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

1. Show that the expectation of $N(\mu, \sigma^2)$ is μ. In fact, σ^2 is the variance of the distribution. This can also be shown using integration by parts, but you don't have to do it right now.
   Solution. By the definition of expectation,

$$\begin{aligned}
\mathbb{E}[X] &= \int_{x=-\infty}^{\infty} x \cdot N(\mu, \sigma^2)(x) dx && (20)\\
&= \int_{x=-\infty}^{\infty} (x - \mu + \mu) \cdot N(\mu, \sigma^2)(x) dx && (21)\\
&= \mu \cdot \int_{x=-\infty}^{\infty} N(\mu, \sigma^2)(x) dx + \int_{x=-\infty}^{\infty} (x-\mu) \cdot N(\mu, \sigma^2)(x) dx && (22)\\
&= \mu + \int_{x=-\infty}^{\infty} (x-\mu) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx && (23)\\
&= \mu + \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{y=-\infty}^{\infty} y \cdot \exp\left(-\frac{y^2}{2\sigma^2}\right) dy && (24)\\
&= \mu, && (25)
\end{aligned}$$

   where we have used that integral of a probability density function is 1, and that $y \exp(-y^2/(2\sigma^2))$ is an odd function hence integrating to 0.

2. Your friend generates X, Y independently from $N(0, 1)$. What is the distribution of $X + Y$, and what is the distribution of $2X$? Assume that they are both Gaussians (so all you need are the means and variances).
   Solution. For any X, Y, we have $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. Hence the mean of $X + Y$ is 0. For independent X, Y, $Variance(X + Y) = Variance(X) + Variance(Y)$, hence variance of $X + Y$ is 2. $2X$ has mean $2\mathbb{E}[X]$ and variance $4Variance(X)$. Hence $X + Y$ and $2X$ are $N(0, 2)$, and $N(0, 4)$ respectively.

3. Your friend then says that you can select one of the following two numbers: $2X$ or $X + Y$. You get a candy if the number you select is larger than 1. What do you choose to maximize your chances of getting the candy?
   Solution. You should choose $2X$. Primarily because of higher variance. Please complete argument here.

Problem 4. 1. Find the point closest to the origin, which is also a solution to:

$$\begin{aligned}
x + y + 2z &= 10\\
x - y + z &= 20.
\end{aligned}$$

Hint: Lagrange Multipliers.

2. Find the point on $x + 2y = 10$, that is closest to the origin.
   Solution. The goal is to minimize $x^2 + y^2$. Substitute $x = 10 - 2y$, and you have to minimize $5(y^2 - 8y + 25)$. This is minimized at $y = 4$, hence the closest point is $(2, 4)$. However, you should do this problem via Lagrange multipliers. That should be more helpful in general settings. The Lagrangian in this case would be:

$$f(x, y, \lambda) = x^2 + y^2 + \lambda(x + 2y - 10).$$

   Then, for the minima:

$$\frac{\partial f(x, y, \lambda)}{\partial \lambda} = 0 \Rightarrow x + 2y = 10, \tag{26}$$

$$\frac{\partial f(x, y, \lambda)}{\partial x} = 0 \Rightarrow 2x + \lambda = 0, \tag{27}$$

$$\frac{\partial f(x, y, \lambda)}{\partial \lambda} = 0 \Rightarrow 2y + 2\lambda = 0. \tag{28}$$

   Solve these equations to obtain $x = 2, y = 4$.

3. The gradient is of a function is its direction of highest increase. What is the gradient of $x^2y + y^2 - xy$ at the point $(1, 1)$?
   Solution. The gradient of f is given by $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial x})$. Plugging the function, we obtain the gradient as $(2xy - y, x^2 + 2y - x)$, which is $(1, 3)$ at $(1, 1)$.

4. What is the smallest and largest value of $x^2 + y^2$ within $(x - 3)^2 + (y - 2)^2 \leq 1$?
   Solution. We want to find the points on and within the circle with center at $(3, 2)$ that are farthest and closest to the origin. These are the intersection points of the line through the origin and $(3, 2)$. The distance of origin to $(3, 2)$ is $\sqrt{13}$. Hence the smallest and largest value of $x^2 + y^2$ is $(\sqrt{13} - 1)^2$, and $(\sqrt{13} + 1)^2$ respectively.

Problem 5. Please do the following. Some links are on the class website, and search online for more.

1. Install Python3.x, and Jupyter on your computer.

2. Install numpy and scipy. We may need more things later.

After this try and play around with python to get a handle of things such as printing, and doing mathematical computations. Feel free to use the internet as much as you like. Watch scipy, and numpy tutorials, and any other videos you like.

Then do the following:

1. Load the file heights-age-test.csv. The file has four test examples, one in each row. Each example has four attributes, $NAME, AGE, BOY?$, and $Height > 55$".

2. You have to implement the decision tree that we did in class in python. In other words, you simply have to do the following:

   (a) Check if the example is a boy.

   (b) If Yes, check if $age > 8$, if No check if $age > 11$.

(c) Output 1 if the answer to the previous point was yes, otherwise 0.

3. Match the answers you get with the last column of the file (Which you should not have touched before!).

4. Submit the code with a read me file that we can use to run the code.

5. The output should be simply one number, the number of test cases that are correctly classified.

The goal of this exercise is more for gaining familiarity with python rather than implementing machine learning algorithms, which we will do in the next assignment.

Solution. See the uploaded python notebook. We have two implementations there, and there could be many more. The difference is in the way the file is read. The first is using csv reader, which reads inputs as strings, and the second is using numpy, which can only read integers. Try to see what happens to the first column of data, when you use numpy.