Briana Zhao
U93987617
EC414
HW 3

1a $P(x|\theta) = \frac{1}{\pi}\left[\frac{1}{(x-\theta)^2+1}\right]$

$$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{1}{\pi}\left[\frac{1}{(x_i-\theta)^2+1}\right]$$

$$= \underset{\theta}{\operatorname{argmax}} \left(\frac{1}{\pi}\right)^n \prod_{i=1}^{n}\left[\frac{1}{(x_i-\theta)^2+1}\right]$$

Taking log: $\underbrace{n\ln\left(\frac{1}{\pi}\right)}_{\text{not dependent on } \theta} + \underbrace{\sum_{i=1}^{n}\ln\left(\frac{1}{(x_i-\theta)^2+1}\right)}$

$$= \sum_{i=1}^{n}\left[-\ln\left((x_i-\theta)^2+1\right)\right]$$

$$= -\sum_{i=1}^{n}\ln\left((x_i-\theta)^2+1\right) = -\sum_{i=1}^{n}\ln\left(x_i^2-2x_i\theta+\theta^2+1\right)$$

deriving: $\frac{\partial}{\partial\theta}\left[-\sum_{i=1}^{n}\ln\left(x_i^2-2x_i\theta+\theta^2+1\right)\right]$

$$= \sum_{i=1}^{n}\frac{-(-2x_i+2\theta)}{x_i^2-2x_i\theta+\theta^2+1} = \sum_{i=1}^{n}\frac{2x_i-2\theta}{x_i^2-2x_i\theta+\theta^2+1}$$

Trying to solve this for θ_{MLE} by setting equal to 0 will not work because it will be very difficult to solve for θ directly after expanding out this sum of polynomials. If n is large, then you will have many terms of polynomials, and solving for θ will be even harder.

1b From part ⓐ, the Log-likelihood was:

$$LL(\theta) = n\ln\left(\frac{1}{\pi}\right) + \sum_{i=1}^{n}\ln\frac{1}{((x_i-\theta)^2+1)}$$

$J(\theta) = -\frac{1}{n}LL(\theta)$ where n is the amount of data, so

$$J(\theta) = -\frac{1}{n}\left[n\ln\left(\frac{1}{\pi}\right) + \sum_{i=1}^{n}\ln\frac{1}{((x_i-\theta)^2+1)}\right]$$

so $\boxed{J(\theta) = -\ln\left(\frac{1}{\pi}\right) - \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1}{(x_i-\theta)^2+1}\right)}$

$\boxed{1c}$ $\frac{\partial}{\partial\theta} J(\theta) = \frac{\partial}{\partial\theta}\left[-\ln\left(\frac{1}{\pi}\right) - \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1}{(x_i-\theta)^2+1}\right)\right]$

$$= -\frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{\partial}{\partial\theta}\ln\left(\frac{1}{(x_i-\theta)^2+1}\right)}_{\text{solved this derivative in part (a)}}$$

$$\frac{\partial}{\partial\theta}J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\frac{2x_i-2\theta}{x_i^2-2x_i\theta+\theta^2+1}$$

$$\theta_{n+1} = \theta_n - \alpha\left[-\frac{1}{n}\sum_{i=1}^{n}\frac{2x_i-2\theta}{x_i^2-2x_i\theta+\theta^2+1}\right]$$

so $\boxed{\theta_{n+1} = \theta_n + \frac{\alpha}{n}\sum_{i=1}^{n}\frac{2x_i-2\theta}{x_i^2-2x_i\theta+\theta^2+1}}$

$\boxed{1d}$ Initiate weight vector randomly
Code in $J(\theta)$, $\frac{\partial}{\partial\theta}J(\theta)$, and the update rule.

for t in [1, ..., T] {
 Set learning rate α
 Use current weight vector, calculate $\frac{\partial}{\partial\theta}J(\theta)$ # need to loop through all training data for this
 Use the update rule to find the next weight vector, θ_{n+1}
 Set weight vector to new weight vector from update rule ($\theta_n := \theta_{n+1}$)
}

Could loop for set # of iterations (1, ..., T) or could loop until $J(\theta)$ is no longer changing (could calculate $J(\theta)$ for each iteration and store in array)

$\boxed{1e}$ Gradient descent update rule from part (c): $\theta_{n+1} = \theta_n + \frac{\alpha}{n}\sum_{i=1}^{n}\frac{2x_i-2\theta}{x_i^2-2x_i\theta+\theta^2+1}$

Want batch size 1! So n=1: $\boxed{\theta_{n+1} = \theta_n + \alpha\left(\frac{2x-2\theta}{x^2-2x\theta+\theta^2+1}\right)}$ [x is chosen randomly]

Initiate weight vector randomly
Code in update rule, $J(\theta)$, and $\frac{\partial}{\partial\theta}J(\theta)$

for t in [1, ..., T] {
 Set the learning rate α

Select a random $\vec{x}$ from data points (mini-batch size of 1)
Use current θ and $\vec{x}$ to compute $\frac{d}{d\theta}J(\theta)$

Use update rule to find next θ (θ_{n+1})
Set $\theta := \theta_{n+1}$

}

#Can loop for T times or can loop until $J(\theta)$ is no longer changing very much

$\boxed{2a}$

P = positive = 'yes'
n = negative = 'no'

Total: $p = 3$, $n = 2$, $p+n = 5$

$$E(S) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.970951$$

Temp Hot: $p=1$, $n=2$, $p+n=3$

$$E(T=hot) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.918296$$

Temp Mild: $p=2$, $n=0$, $p+n=2$

$$E(T=mild) = -\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right) = 0$$

$$I(Temp) = (0.918296)\left(\frac{3}{5}\right) + 0\left(\frac{2}{5}\right) = 0.550978$$

$$Gain(Temp) = 0.970951 - 0.550978 = 0.419973 \star$$

Humidity Normal: $p=2$, $n=2$, $p+n=4$

$$E(H=normal) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$

Humidity High: $p=1$, $n=0$, $p+n=1$

$$E(H=high) = -\frac{1}{1}\log_2(1) - \frac{0}{1}\log_2\left(\frac{0}{1}\right) = 0$$

$$I(Humidity) = 1\left(\frac{4}{5}\right) + 0\left(\frac{1}{5}\right) = \frac{4}{5} = 0.8$$

$$Gain(Humidity) = 0.970951 - 0.8 = 0.170951 \star$$

Wind Strong: $p=2$, $n=2$, $p+n=4$

$$E(W=Strong) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = 1$$

Wind weak: $p=1$, $n=0$, $p+n=1$

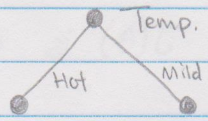
$$E(W=weak) = -\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right) = 0$$

$$I(\text{Wind}) = 1\left(\frac{4}{5}\right) + 0\left(\frac{1}{5}\right) = \frac{4}{5} = 0.8$$

$$\text{Gain}(\text{wind}) = 0.970951 - 0.8 = \underline{0.170951}$$ ☆

Temperature had the largest information gain (0.419973), so the $\boxed{\text{Root node is Temperature.}}$

$\boxed{2b}$



Repeating 2a steps, but now only using 'Hot' data

<u>Total hot:</u> $P=1,\ n=2,\ p+n=3$

$$E(S) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) = 0.918296$$

<u>Humidity Normal:</u> $P=1,\ n=2,\ p+n=3$

$$E(H=\text{normal}) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.918296$$

<u>Humidity High:</u> $P=0,\ n=0,\ P+n=0$

$$E(H=\text{high}) = 0$$

$$I(\text{Humidity}) = 0.918296\left(\frac{3}{3}\right) + 0\left(\frac{0}{3}\right) = 0.918296$$

$$\text{Gain}(\text{Humidity}) = 0.918296 - 0.918296 = \underline{0}$$ ☆

<u>Wind Strong:</u> $P=0,\ n=2,\ p+n=2$

$$E(W=\text{strong}) = -\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right) = 0$$

<u>Wind Weak:</u> $P=1,\ n=0,\ p+n=1$

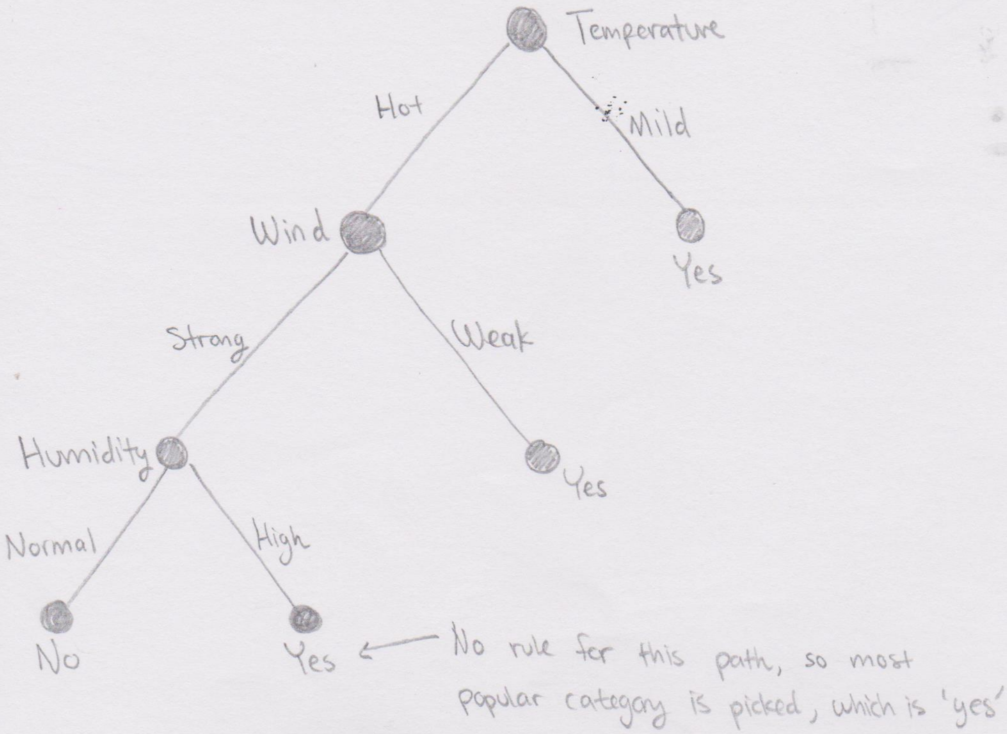
$$E(W=\text{weak}) = -\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right) = 0$$

$$I(\text{Wind}) = 0\left(\frac{2}{3}\right) + 0\left(\frac{1}{3}\right) = 0$$

$$\text{Gain}(\text{Wind}) = 0.918296 - 0 = \underline{0.918296}$$ ☆

Gain is highest with Wind so the next node is $\boxed{\text{Wind}}$

The last node will be humidity because it is the remaining attribute.


Temperature
Hot
Mild
Wind
Yes
Strong
Weak
Humidity
Yes
Normal
High
No
Yes
No rule for this path, so most popular category is picked, which is 'yes'