Hw 7

Neel Singh

12/3/2024

Recall that in class we showed that for randomized response differential privacy based on a fair coin (that is a coin that lands heads up with probability 0.5), the estimated proportion of incriminating observations $\hat{P}$ [1] was given by $\hat{P} = 2\hat{\pi} - \frac{1}{2}$ where $\hat{\pi}$ is the proportion of people answering affirmative to the incriminating question.

I want you to generalize this result for a potentially biased coin. That is, for a differentially private mechanism that uses a coin landing heads up with probability $0 \leq \theta \leq 1$, find an estimate $\hat{P}$ for the proportion of incriminating observations. This expression should be in terms of θ and $\hat{\pi}$.

Response:

The observed proportion of positive responses $\hat{\pi}$ is a mix of truthful responses and random responses. The proportion of truthful and positive responses is $\hat{P}$, and a proportion θ participants have the opportunity to tell a truthful response (based on an initial coin flip). The proportion of random and positive responses is 0.5, and a proportion of $1 - \theta$ participants have the opportunity to tell a random response. Thus, the expected proportion of positive responses $\hat{\pi}$ is $\theta\hat{P} + (1 - \theta)0.5$.

$$\hat{\pi} = \theta\hat{P} + \frac{1 - \theta}{2}$$

$$\hat{\pi} = \theta\hat{P} + \frac{1}{2} - \frac{\theta}{2}$$

$$\hat{\pi} - \frac{1}{2} + \frac{\theta}{2} = \theta\hat{P}$$

$$\hat{P} = \frac{\hat{\pi} - \frac{1}{2} + \frac{\theta}{2}}{\theta}$$

$$\hat{P} = \frac{\hat{\pi} - \frac{1}{2}}{\theta} + \frac{1}{2}$$

[1] in class this was the estimated proportion of students having actually cheated

Next, show that this expression reduces to our result from class in the special case where $\theta = \frac{1}{2}$.

$$\hat{P} = \frac{\hat{\pi} - \frac{1}{2}}{\frac{1}{2}} + \frac{1}{2}$$

$$\hat{P} = \frac{\hat{\pi} - \frac{1}{2}}{\frac{1}{2}} + \frac{1}{2}$$

$$\hat{P} = \frac{\hat{\pi}}{\frac{1}{2}} - \frac{\frac{1}{2}}{\frac{1}{2}} + \frac{1}{2}$$

$$\hat{P} = 2\hat{\pi} - 1 + \frac{1}{2}$$

$$\hat{P} = 2\hat{\pi} - \frac{1}{2}$$

Part of having an explainable model is being able to implement the algorithm from scratch. Let's try and do this with `KNN`. Write a function entitled `chebychev` that takes in two vectors and outputs the Chebychev or L^∞ distance between said vectors. I will test your function on two vectors below. Then, write a `nearest_neighbors` function that finds the user specified k nearest neighbors according to a user specified distance function (in this case L^∞) to a user specified data point observation.

```
#student input
#chebychev function
chebychev <- function(vec1, vec2) {
  max(abs(vec1 - vec2))
}

nearest_neighbors = function(x, obs, k, dist_func){
  dist = apply(x, 1, dist_func, obs)
  distances = sort(dist ) [1: k]
  neighbor_list = which(dist %in% sort(dist)[1:k])
  return( list (neighbor_list, distances))
  }

x<- c(3,4,5)
y<-c(7,10,1)
chebychev(x,y)
```

```
## [1] 6
```

Finally create a `knn_classifier` function that takes the nearest neighbors specified from the above functions and assigns a class label based on the mode class label within these nearest neighbors. I will then test your functions by finding the five nearest neighbors to the very last observation in the `iris` dataset according to the `chebychev` distance and classifying this function accordingly.

```
library(class)
df <- data(iris)
#student input
knn_classifier = function(x,y){
groups = table(x[,y])
pred = groups[groups == max(groups)]
return(pred)
}


#data less last observation
x = iris[1:(nrow(iris)-1),]
#observation to be classified
obs = iris[nrow(iris),]

#find nearest neighbors
ind = nearest_neighbors(x[,1:4], obs[,1:4],5, chebychev)[[1]]
as.matrix(x[ind,1:4])
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width
## 71           5.9         3.2          4.8         1.8
## 84           6.0         2.7          5.1         1.6
## 102          5.8         2.7          5.1         1.9
## 127          6.2         2.8          4.8         1.8
## 128          6.1         3.0          4.9         1.8
## 139          6.0         3.0          4.8         1.8
## 143          5.8         2.7          5.1         1.9
```

```
obs[,1:4]
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width
## 150          5.9           3          5.1         1.8
```

```
knn_classifier(x[ind,], 'Species')
```

```
## virginica
##         5
```

```
obs[,'Species']
```

```
## [1] virginica
## Levels: setosa versicolor virginica
```

Interpret this output. Did you get the correct classification? Also, if you specified $K = 5$, why do you have 7 observations included in the output dataframe?

Response:

Based on the $K = 5$ nearest neighbors, the observation belongs to the virginica class. This classification was correct, aligning with the true class label. I got 7 observations despite specifying $K = 5$ likely because there were a few ties in Chebychev distance between the observation and neighboring points, causing the `nearest_neighbors` function to report the seven nearest neighbors.

Earlier in this unit we learned about Google's DeepMind assisting in the management of acute kidney injury. Assistance in the health care sector is always welcome, particularly if it benefits the well-being of the patient. Even so, algorithmic assistance necessitates the acquisition and retention of sensitive health care data. With this in mind, who should be privy to this sensitive information? In particular, is data transfer allowed if the company managing the software is subsumed? Should the data be made available to insurance companies who could use this to better calibrate their actuarial risk but also deny care? Stake a position and defend it using principles discussed from the class.

Response:

The harm principle asserts that individuals or institutions are justified in limiting someone's freedom of action only to prevent harm to others. Conversely, individuals or institutions should not be limited in freedom unless to prevent harm to others. From the perspective of the harm principle, algorithmic assistance in healthcare is permissible as long as its use of sensitive data does not cause harm. If patient data is employed solely to improve medical outcomes and enhance patient care, no harm is being inflicted, and thus there is no ethical reason to prevent its use. However, granting access to this data to entities like insurance companies, who may use it to deny coverage or increase rates for patients whose data it acquires, introduces the potential for significant harm. This misuse of data would violate the harm principle, as the freedom of insurance companies to use this data would directly harm individuals. Therefore, while the use of data for algorithmic healthcare assistance to improve patient outcomes is justified, strict safeguards must ensure that access to sensitive data must not be granted for purposes that can do harm. The legal system must ensure that patient data remains separate from potential harmdoers in the event of the current caretaker of patient data being subsumed.

I have described our responsibility to proper interpretation as an *obligation* or *duty*. How might a Kantian Deontologist defend such a claim?

Response

In statistical or ML contexts, proper interpretation ensures the responsible use of data, models, and outcomes, avoiding misrepresentation or misuse that could lead to harm or injustice (like COMPAS). Kantian Deontology is an ethical framework based on the idea that moral actions are determined by duty. It asserts that every individual must never be treated merely as a means to an end. This means that people should not be used solely to achieve a goal or serve another's purpose without consideration of their inherent dignity and autonomy. In the context of statistical or machine learning interpretation, a Kantian Deontologist would argue that proper interpretation is a moral duty because it ensures that data and results are used responsibly and ethically. Misinterpretation, whether through negligence or intentional distortion, could lead to decisions that reduce the humanity of individuals based solely on incorrect interpretation of algorithms,

violating their dignity. The results from algorithms like COMPAS can easily be misinterpreted, perhaps by individuals who confuse a risk assessment tool for a crystal ball. This could strip the dignity and humanity of an individual without good reason. This underscores the Kantian imperative to approach statistical and ML interpretation with rigor and care, as failing to do so compromises the dignity of individuals affected by these decisions.