Hw 7

Hunter Busick

12/2/2024

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 $\hat{\pi}$.

$\hat{\pi}$ consists of those who answer truthfully to a heads observation P, and $1 - P$ for those who answer "yes" when the coin lands on tails. We can express this as

$\hat{\pi} = \theta P + ((1 - \theta)(1 - P))$

Expanding the last term we get

$\hat{\pi} = \theta P + 1 - \theta - (1 - \theta)P$

Combining terms

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

Arranging for P in terms of $\hat{\pi}$

$\hat{\pi} - (1 - \theta) = (2\theta - 1)P$

Leaving us with

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

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

From before we know that $\hat{\pi} = \theta P + (1 - \theta)1$

We can then substitute in $\frac{1}{2}$ for θ and get

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

Solving for P

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

Solving out

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

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

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.

```
cheby <- function(x,y) {
  return(max(abs(x-y))) #This is just the maximum absolute difference between the vectors
}

nearest_neighbors <- function(x, obs, k, dist_fun) {
  #apply distance function accross all rows of x
  distances <- apply(x, 1, function(row) dist_fun(row, obs))
  #assign distances
  ranks <- rank(distances)
  #use which to determine which distances to keep <= k
  nearest_index <- which(ranks <= k)
  return(list(indices = nearest_index, neighbors = x[nearest_index, , drop = FALSE]))
}

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

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)

knn_classifier <- function(neighbors, class_column) {
  #get class labels from all rows
  class_lbl <- neighbors[, class_column]
  #sort a table in decreasing order of the labels of closest observations. Grab the first
  label <- names(sort(table(class_lbl), decreasing = TRUE))[1]
  return(label)
}

#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, cheby)[[1]]
as.matrix(x[ind,1:4])
```

```
obs[,1:4]
knn_classifier(x[ind,], 'Species')
obs[,'Species']
```

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?

We see that we do get the correct classification here as the 150th observation which we removed is a virginica flower, which is what we correctly classify it as based on the observations closest to it. Despite specifying $K = 5$ we have 7 neighbors considered in our output dataframe. This is because in constructing the classifier we did not put any method for handling ties. Thus we must have flowers that have the same distance from our chosen flower, but rather than flipping a coin to decide which is considered, we have simply included all ties.

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.

Appealing to the framework laid out by deontology, it is not acceptable for data transfer to be used for corporate gain, or for data to be seamlessly transferred over when a company managing it is taken over by another. This is especially important pertaining to healthcare data. One of our core formulations is that we need to be able to universalize the act of transferring data or giving it to other companies for it to pass Kant's categorical imperative. We cannot universalize patients having their data given to others because their is often a tacit relationship between a doctor and their patients. Every time you visit the doctor they do not give you the rundown about how they won't sell your data to another company, it is simply implied. We would need individual consent from each patient confirming that they are okay with their data being shared with universities for instance to better help researchers. It would take explicit consent in this instance to violate the categorical imperative. Additionally, the sole purpose of an insurance company purchasing medical records is to try and perfectly train their models such that they offer (what they consider to be) the perfect premium for each client. This a textbook definition of using someone as means to an end, our second formulation. Regardless of whether or not a client receives a better or worse premium, the company will be able to better maximize their profits at the expense of those who are deemed to be "too risky" for cheaper premiums. This would result in a net loss of utility in the long run, appealing to consequentialism that is.

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

A Kantian Deontologist would defend the claim that we have an obligation to proper interpretation by arguing that we violate the categorical imperative if we were not to do so. We can think of not interpreting our methods or results as what we are trying to universalize. While there may be instances in which certain

methods and algorithms should remain protected (national security), we cannot claim that there is a right to privacy for all statistical work. As a prime example, one of the largest grievances of the COMPAS algorithm is that it is black-boxed, thus not allowing the public to critique and examine it. While we cannot make the claim that all algorithms should be open source, this does argue that it should be the standard we strive for. Secondly, we must consider that we use others as means to an end when we fail to properly interpret methods and results. A salient example of this can be found in a graphic displayed by many news networks following the election highlighting a stark red shift in states like NY and California. Without mentioning the fact that this was primarily due to voter apathy in one sided regions, viewers are treated as means to an end as networks seek to shape their opinion assisted by results that are poorly interpreted.