Hw 7

Holden Wright

1/5/2024

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 $\hat{\pi}$.

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

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

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

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

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

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

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

cheby <- function(vec1, vec2) {
  max(abs(vec1 - vec2))
```

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

```
}

#nearest_neighbors function

nearest_neighbors = function(x, obs, k, dist_func){
dist = apply(x, 1, dist_func, obs) #apply along the rows
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)
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)
#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, 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 do get the correct classification with virginica. There are 7 observations included in the output because there are ties according to the chebychev distance.

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.

Google's DeepMind offers significant potential for improving patient outcomes. However, the acquisition, retention, and use of sensitive healthcare data raises ethical concerns. My position is that healthcare data should only be accessible to the parties directly involved in patient care or research that benefits public health, and data transfer should be heavily restricted even outlawed, in cases of corporate mergers or acquisitions. Further, sharing data with insurance companies for actuarial purposes, should be prohibited unless patients provide explicit, informed consent.

Patients deserve autonomy over their own personal data and third parties such as insurance companies do not have a claim to this data. Transferring data to insurance companies has potential to create harm by companies using the data to create discriminatory practices. In the case of Google's DeepMind the potential to benefit the patient outweighs potential privacy concerns. However, should another company acquire the software the data should not be transferred over. This creates too large of a data leakage risk and infringes on the patients privacy.

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

Proper interpretation is grounded in respecting the rationality and autonomy of the people whose ideas or actions are being interpreted. Misinterpretation undermines their ability to be understood as rational agents, effectively treating their expressed thoughts or intentions as means to our end. A Kantian would argue that proper interpretation is a duty we owe to others because it honors their rationality and intrinsic dignity.