Hw 7

COLLIN REGISTER

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

Student Input Originally, we would get the expression $\pi = \theta(\hat{P}) + (1-\theta)\theta$. That expression simplifies down to $\hat{P} = \frac{(\pi-(1-\theta)\theta)}{\theta}$ when we rearrange to have $\hat{P}$ by itself.This makes sense because theta times 1-theta represents the probability of tails on the first coin flip and saying yes for the second one and theta times p represnats the probability of heads on the first coin flip.

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

Student Input This expression reduces down to the results from class by the following steps. If we replace theta with 1/2 in the expression $\pi = \theta(\hat{P}) + (1-\theta)\theta$ and then rearrange for to get p by itself we will get p=(pi-1/4)2 which simplifies to $\hat{P} = 2\pi - \frac{1}{2}$

Consider the additive feature attribution model: $g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i$ where we are aiming to explain prediction f with model g around input x with simplified input x'. Moreover, M is the number of input features.

Give an expression for the explanation model g in the case where all attributes are meaningless, and interpret this expression. Secondly, give an expression for the relative contribution of feature i to the explanation model.

Student Input If all of the attributes are meaningless, the the expression would then be just $g(x') = \phi_0$ because the there are not input features to calculate for. This expression would just be the prediction regardless of inputs meaning it is a constant set at ϕ_0. The relative contribution of the feature i to the explanation model is i= $\phi_i/(\sum_{i=1}^{M} \phi_i x'_i)$. This helps us see the proportion of values that can be attributed to i.

[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.

```
#chebychev function
cheby <- function(x, y) {
  max(abs(x - y))}

#nearest_neighbors function
nearest_neighbors= function(x,abs,k,dist_func){
  dist=apply(x,1,dist_func,abs)
  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, chebychev)[[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?

Student Input This output is taking the five closest points according to the chebychev distance and using it to classify the point based on those values and their classifications. I did get the correct classification, which is Virginica. Observation 102 and 139 have the same value so that is why there is an additional row and then the observation itself is also including which gives us the 7 observations.

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.

Student Input The sensitive information should only be allowed to be accessed by those who collected the data with consent. Health insurance companies and others should not be allowed access and the data should be held safely with those who originally acquired it. If access were given to other companies, this would lead to a slippery slope and informed consent would not mean anything because people would not be fully informed on where their data was being sent to. I don't think the data should be transferred if the company is subsumed because the initial informed consent given, has now changed and the new company may have new processes/regulations for handling the data that the patient should now be made aware of. An insurance company should not be allowed data to calculate risk to deny healthcare because this could now invoke the harm principle. With this, the use of this data could now be used against those who gave access and the company should not longer have to right to do what they want with this data/ give it away to insurance companies.