STOR390HW7

Jillian Myler

2024-04-18

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.

$\pi = \theta(\hat{P}) + (1-\theta)\theta$ Hence, solving for $\hat{P}$ in terms of θ and π

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

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

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

Thus, when $\theta = \frac{1}{2}$,

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

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

$\hat{p} = \frac{\pi-\frac{1}{4}}{\frac{1}{2}}$

$\hat{p} = 2\pi - \frac{2}{4}$

which reduces 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.

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

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.

Expression for all attributes being meaningless: $\frac{|S|!(|F|-(|F|-1)-1)!}{|F|!} = \frac{|S|!(0!)}{|F|!} = \frac{1}{|F|}$

expression for relative contribution of feature i to the explanation model: $S = j \neq i$

$|S| = 1$ $\frac{1!(|F|-2)!}{|F|!}$ $\frac{(|F|-2)!}{|F|!}$ $\frac{1}{|F|*|F-1|}$

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

```
## [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)

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])
```

```
##     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?

This output means that observation 71 is the most similar to our selected observation (150). Further, based on the species of these 5 nearest neighbors being mostly virginica, the predicted species type of our observation is virginica. In fact, the true species of our observation is virginica so our classification is correct. To discuss why there are 7 outputs when the k specified was 5, given the way that the chebychev distance measure works: x,y: d(x)= $max|x_i, y_i|$, it is possible for the distances calculated between the selected observation and another observation to be the same as the distance from that selected observation and another.In other words, there are 7 observations included in the output dataframe instead of 5 because there are multiple observations yeilding the same distance metric from the selected observation, i.e. the distances are tied.

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.

Indeed assistance in the health care sector is always welcome, especially when it benefits the well being of a patient, and especially if it benefits many patients. However, when thinking about the data that goes into training an algorithm to help alleviate medical problems, it is clear that there is sensitive information present for each person in the data set. One thing discussed in this class that should be standard is that the data should be protected to the point where the individual should not be able to be traced back to their data. The data should be anonymized well. When there are cases that may turn out to be an outlier, maybe a complicated case where there turns out to be something deeper going on medically than originally perceived that may be used in an update of the algorithm in the future, it should still remain standard that the person is only identifiable as an id code in a similar fashion to how Jane/John Doe are names that are used in medical case descriptions. I think there is value to being able to identify a particular point within an algorithm to update the algorithm in its entirety to be better, but that person should still remain anonymous. Therefore, I think the only people who should know the exact information of a patient within the dataset should be the doctor directly involved in helping them, otherwise, it should be anonymized completely. This reasoning would follow in line with how the healthcare system already requires doctor patient confidentiality and release forms for sharing of medical information across teams/external providers. Thus, in regards to a company subsuming another, there should be an updated contract with those who's data is in use for the new company to continue to use it. I think there being transparency as to who has someone's sensitive data is imperative in order to continue to respect informed consent. Further, I do not believe the insurance companies should be able to use the information to better their risk assessment if it comes at the ability to deny coverage for individuals who need it. I think at the consequentialist approach, allowing an insurance company to assess risk to the extent that they start denying coverage and reducing access to those with higher risk would allow for a healthcare system to operate in a way that is no longer aiming to serve those who need it the most because they likely then would not be able to afford care; the marginal benefit to an insurance company would be far outweighed by the aggregate harm to the population, where care resources are already limited, and would cause denial of care to those who need it most directly violating the difference principle.