Hw 7

Ethan Turner

4/12/2024

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 Answer

From the perspective of π, π now = $\left(\theta * \hat{P}\right) + (1-\theta)(\theta)$

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

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

Student Answer

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

Substituting $\frac{1}{2} = \theta$:

$$\frac{\pi}{\frac{1}{2}} - \left(1 - \frac{1}{2}\right) = 2\pi - \frac{1}{2}$$

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

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 Answer

First, where all attributes are meaningless: No values contribute anything to the prediction. Thus, we have g(x’)=ϕ_0. Without the input features, our model prediction is this constant term ϕ_0.

Second, an expression for relative contribution of feature *i*:

I want to weigh each individual feature in comparison to all features, so I would compute:

$\frac{\phi_i x_i{}'}{\sum_{j=1}^{M} \phi_j x_j{}'}$. I shifted the notation of the denominator to *j* to distinguish the components of the sum of the model features from the individual feature *i*.

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.

```r
#student input
#chebychev function
cheby <- function(x1, y1){
  return(max(abs(x1-y1)))
}


#nearest_neighbors function

nearest_neighbors <- function(group, single, k, kfunc) {

  distances <- apply(group, 1, function(input) cheby(input, single))
  neighbors <- group[distances <= sort(distances)[k], ]

  return(neighbors)
}


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)
library(tidyverse)

df <- data(iris)
#student input

knn_classifier <- function(neighbors, label){

 neighbor_labels <- neighbors[[label]]

  count <- table(neighbor_labels)

  majority_class <- names(count)[which.max(count)]

  return(majority_class)
}


#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
## 5            5.0         3.6          1.4         0.2
## 6            5.4         3.9          1.7         0.4
## 5.1          5.0         3.6          1.4         0.2
## 6.1          5.4         3.9          1.7         0.4
## 6.2          5.4         3.9          1.7         0.4
## 6.3          5.4         3.9          1.7         0.4
## 5.2          5.0         3.6          1.4         0.2

obs[,1:4]

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

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

## [1] "setosa"

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?

The classification is incorrect. We classified the observation as setosa when it was really virginica. We have 7 observations because we must have had multiple (at least three) observations of the same Chevyshev distance at the upper threshold/cutoff of selection as 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.

Student Answer

Firstly, I do not believe that the subsumation of companies should allow third parties to gain access to sensitive health care data. When a patient provides fully informed consent before agreeing to share their data in the first place, they are assumedly unaware at the time of consent that their data will change hands in the future. Unless this point is made clear to a patient at the time of data collection, which is surely unlikely to be common practice, I struggle to morally rationalize this approach. Further, in this individual case of applying data to insurance and potentially depriving care based on conclusions drawn from transferred data, I also postulate that individuals ought to possess the autonomy to declare whether their personal data may be used at the cost of others. Hence, the initial consent to provide data does not also cover a transfer to other lines of exploration.

Regardless of purpose, data transfer without explicit informed consent for insurance companies and so forth raises severe concerns. Although this specific case of data transfer may not inherently violate the Harm Principle (essentially that actions should be limited when they cause harm to others), the allocation of rights to corporate bodies to transfer data however they wish can clearly lead to violations. Without checks to corporate power, shadowy organizations can easily manipulate individual data for nefarious means to extort customers into overpaying based on calculated pressure points or deprive service based on profile. These profiles, depending on the model's biases, may also fall on disturbing social lines as seen by the application of the COMPAS algorithm (certainly shadowy and mysterious if not inherently a product of data transfer) along racial boundaries. Thus, I argue that we ought to avoid the very danger of setting such a precedent for data transfer.

Acquired personal data may indeed spur on many medical advancements, but to the broad question of who should possess access to helpful personal data I feel that we must prioritize personal autonomy and agency. In other words, directly aiding medical research is not a ubiquitous obligation. Explaining the utility of personal data in a request to individuals to apply their own individual data may well sway throngs to volunteer for medical studies, but the individual informed decision of each potential study participant still takes the highest priority. Therefore, we sacrifice time and convenience for our normative aims.