Hw 7

Seth Hall

11/30/2024

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 $hatP$ [1] was given by $hatP = 2hat\pi - \frac{1}{2}$ 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 $hatP$ for the proportion of incriminating observations. This expression should be in terms of θ and $hat\pi$.

If the coin lands heads up with probability theta, the student tells the truth. If the coin lands tails up with probability 1 - theta, the respondent flips the coin again and answers yes with probability theta and no with probability 1 - theta. The probability of answering yes is then expressed as:

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

where P is the actual proportion of incriminating observations. Solving for P, we get:

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

$$\hat{\pi} = \theta P + \theta - \theta^2$$

$$\hat{P} = \frac{\hat{\pi} - \theta + \theta^2}{\theta}$$

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

Substitute $\theta = \frac{1}{2}$ into the formula:

$$\hat{P} = \frac{\hat{\pi} - \frac{1}{2} + \left(\frac{1}{2}\right)^2}{\frac{1}{2}}$$

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

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

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

This matches the result from class.

[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.

```
#student input
cheby <- function(x, y) {
  return(max(abs(x - y)))
}
nearest_neighbors <- function(data, obs, k, dist_func) {
  distances <- apply(data, 1, function(row) dist_func(row, obs))
  neighbor_indices <- order(distances)[1:k]
  return(list(neighbor_indices, distances[neighbor_indices]))
}


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(neighbors, class_col) {
  classes <- neighbors[, class_col]
  return(names(sort(table(classes), decreasing = TRUE))[1])
}


x <- iris[1:(nrow(iris) - 1), ]
x = iris[1:(nrow(iris)-1),]
obs <- iris[nrow(iris), ]
obs = iris[nrow(iris),]

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']
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?

I did get the correct classification because the predicted classes match the actual class of the last observations in iris. For example, my predicted class for sepal length is 5.9 and the last observation is 5.9. The reason 7 observations are included in the output dataframe could be because multiple observations have the same distance to the target observation.

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.

Obviously the patients themselves should be privy to their own sensitive information. However, it is important that a few other parties are also privy to it to provide accurate diagnoses, treatment plans, and ongoing care. Sensitive healthcare data should primarily be accessible to healthcare providers (doctors/surgeons, nurses) and authorized healthcare entities to ensure the right kind of care. When a company managing such data is acquired, it's important that the patient gives explicit and informed consent and that robust data protection measures are taken to maintain trust and comply with regulations. Insurance companies' access to health data can help in risk assessment and fraud detection but it also raises privacy concerns and the potential for discrimination (e.g. the issue described in my midterm project.) Making sensitive healthcare data available to insurance companies for better actuarial risk calibration raises a lot of ethical concerns. From a consequentialist perspective, while it might lead to more accurate risk assessments, it could also result in denying care to high-risk individuals, increasing their suffering and worsening health outcomes. Deontologically, using data to deny care treats individuals as means to an end, violating their dignity and autonomy. Virtue ethics emphasizes compassion and justice, which are compromised when care is denied based on data analytics. Rawlsian justice, which prioritizes fairness and the protection of the least advantaged, is also at odds with this practice, as it could exacerbate inequalities. Autonomy and informed consent are crucial; sharing data without explicit consent undermines personal autonomy and can lead to coercive practices. Additionally, there is a risk of algorithmic bias, which could result in unfair treatment of certain groups. Overall, the ethical implications of denying care based on data analytics outweigh the benefits of improved risk assessment, so such data should not be made available to insurance companies.

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 argue that proper interpretation is a duty because it respects the dignity and rationality of all of the moral agents involved. Misinterpretation or misleading information would treat individuals as mere means to an end. Proper interpretation ensures transparency, honesty, and respect for others and the autonomy they have, which aligns with the principles of deontology.