Hw 7

UMA NAIR

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 $\hat{\pi}$.

Given a biased coin that lands heads with probability θ, the proportion of people who answer affirmatively to the incriminating question is $\hat{\pi}$. The true proportion of incriminating observations, P, can be estimated by:

$$\hat{\pi} = \theta P + (1-\theta) \times 0.5$$

Solving for P:

$$P = \frac{\hat{\pi} - 0.5(1-\theta)}{\theta}$$

Thus, the estimate for the proportion of incriminating observations, $\hat{P}$, based on a biased coin with probability θ of landing heads is:

$$\hat{P} = \frac{\hat{\pi} - 0.5(1-\theta)}{\theta}$$

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

Given the formula for the estimated proportion of incriminating observations, $\hat{P}$, with a biased coin:

$$\hat{P} = \frac{\hat{\pi} - 0.5(1-\theta)}{\theta}$$

We want to show that this formula reduces to the result from class when $\theta = \frac{1}{2}$.

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

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

$$\hat{P} = \frac{\hat{\pi} - 0.5(1 - \frac{1}{2})}{\frac{1}{2}}$$

Simplifying the terms inside the parentheses:

$$\hat{P} = \frac{\hat{\pi} - 0.5 \times \frac{1}{2}}{\frac{1}{2}}$$

$$\hat{P} = \frac{\hat{\pi} - 0.25}{\frac{1}{2}}$$

Now, dividing by $\frac{1}{2}$ is the same as multiplying by 2:

$$\hat{P} = 2\left(\hat{\pi} - 0.25\right)$$

Distributing the 2:

$$\hat{P} = 2\hat{\pi} - 0.5$$

Thus, we have shown that when $\theta = \frac{1}{2}$, the formula reduces to:

$$\hat{P} = 2\hat{\pi} - 0.5$$

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(x, y) {
  # making sure both vectors are of the same length
  if(length(x) != length(y)) {
    stop("Vectors must have the same length")
  }

  # calculating the Chebychev distance
  distance <- max(abs(x - y))
  return(distance)
}

x<- c(3,4,5)
y<-c(7,10,1)
cheby(x,y)
```

```r
#nearest_neighbors function

nearest_neighbors <- function(data, target, k, dist_func) {
  # empty vector to store distances
  distances <- numeric(nrow(data))

  # calculating the distance from the target to each data point (from previous homeworks)
  for (i in 1:nrow(data)) {
    distances[i] <- dist_func(target, data[i, ])
  }

  # sorting the distances and get the indices of the k smallest distances
  sorted_indices <- order(distances)
  nearest_indices <- sorted_indices[1:k]

  # return the k nearest neighbors (the rows of 'data' corresponding to the k nearest distances)
  return(data[nearest_indices, , drop = FALSE])
}

# two given vectors
x <- c(3, 4, 5)
y <- c(7, 10, 1)

# Chebychev distance function
cheby(x, y)

# creating small dataframe to test (rows are observations, columns are features)
data <- matrix(c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), nrow = 4, byrow = TRUE)

# Define the target observation
target <- c(5, 6, 7)

# Find the 2 nearest neighbors using the Chebychev distance
nearest_neighbors(data, target, k = 2, dist_func = cheby)
```

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.

```r
library(class)
df <- data(iris)
#student input

# KNN Classifier Function
knn_classifier <- function(nearest_neighbors_data, class_column_name) {
  # extracting the class labels ('Species') from the nearest neighbors data
  class_labels <- nearest_neighbors_data[, class_column_name]
  # finding most frequent class in the nearest neighbors
  mode_class <- names(sort(table(class_labels), decreasing = TRUE))[1]

  return(mode_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, cheby)[[1]]
as.matrix(x[ind,1:4])
obs[,1:4]
knn_classifier(x[ind,], 'Species')
obs[,'Species']

print(ind) #to make sure I selected five rows...

# more double checking to see if I did the classification correctly!
predicted_class <- knn_classifier(x[ind, ], 'Species')
actual_class <- obs[,'Species']

print(predicted_class)  # predicted class
print(actual_class)     # actual class
```

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 am seeing 5 rows of neighbors in your output, but only one row is being classified, this indicates that the mode (the most frequent class label) among those 5 neighbors is being returned correctly. In KNN algorithms, if multiple neighbors have the same distance (e.g., the 5th and 6th nearest neighbors are at the same distance), the function may return more than the requested number of neighbors, leading to "extra" observations being included.The 7 observations part of the question may refer to ties in distances, but the algorithm correctly selects the most frequent class from the nearest neighbors, hence returning only one class label.

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.

The ethical concerns surrounding the use of sensitive health care data in the context of AI tools like Google's DeepMind require careful consideration of privacy, consent, and the potential consequences for patients. Health care data should not be made available to insurance companies, and that strict controls should govern who has access to sensitive health data, particularly in cases of corporate acquisition or data transfer. This position is grounded in the principles of privacy, autonomy, and justice, which are central to medical ethics and data privacy discussions. The privacy and confidentiality of patient data are important principles in healthcare ethics. Patients trust medical providers to safeguard their personal information, particularly sensitive data about their health conditions, treatments, and medical history. When companies like DeepMind handle this data to assist in managing acute kidney injury or other conditions, they must adhere to strict standards of privacy protection. If that data is transferred or used for purposes beyond direct care—such as being shared with insurance companies, it could undermine this trust. If data were made available to insurance companies, it could be used to discriminate against patients, potentially leading to denial of coverage or higher premiums for those deemed to be at higher risk of certain conditions. This violates the ethical principle of privacy, as it exposes individuals to potential harms that are unrelated to the primary purpose of data collection, which is the improvement of patient care. Informed consent is another big issue when discussing data privacy. The principle of autonomy asserts that individuals have the right to make informed decisions about their personal data, including how it is used, who has access to it, and for what purposes. This means patients should be fully informed about how their data will be handled, with the option to provide or withhold consent for data sharing.In the case of AI-based tools like DeepMind, if the company managing the software is acquired or merged with another entity, this raises questions about whether patients' consent for data use extends to the new company. Data retention and transfer across corporate boundaries without clear, explicit patient consent can be seen as a violation of autonomy, as individuals may not have been fully informed about the new uses to which their data may be put. Moreover, the principle of informed consent requires that patients have a say in whether their data can be shared with third parties, especially in situations where there is a risk of harm, such as the potential for insurance companies to use data for purposes that could disadvantage the patient. Furthermore, When a company like DeepMind is acquired or merged with another entity, the question of what happens to the patient data it holds becomes crucial. Under most data protection laws, such as Health Insurance Portability and Accountability Act (HIPAA), patient data must be handled with the same level of protection and transparency, even in the case of corporate acquisition. Lastly, The principle of justice is concerned with fairness and the equitable distribution of benefits and burdens. Making sensitive health data available to insurance companies could disproportionately affect vulnerable populations. For instance, individuals who are already suffering from acute kidney injury or other serious health conditions may be unfairly penalized by insurance companies that have access to this information.This could lead to a social justice issue, where individuals are denied care or face higher costs based on predictive algorithms that may not always account for the nuances of a person's medical history or personal circumstances. In conclusion, health care data used for AI-assisted diagnosis and treatment should not be shared with insurance companies. Any data sharing, especially in the case of mergers and acquisitions, should be done with the explicit, informed consent of the patient, and any third-party use of the data should be tightly regulated to avoid abuse.

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 defend the claim that we have a responsibility to proper interpretation of data or actions as a moral obligation or duty by appealing to Kant's principles

of moral duty, particularly the categorical imperative and the idea of respect for persons. Immanuel Kant's ethical theory is grounded in the categorical imperative, a central concept in deontological ethics. The categorical imperative asserts that an action is morally right if it can be universalized—that is, if it can be consistently willed as a law that everyone could follow, without contradiction. According to this principle, we must treat others in a way that respects their inherent dignity and autonomy, and not merely as a means to an end. In the context of proper interpretation, a Kantian would argue that we have an obligation to interpret information—whether it be data, actions, or communications—in a way that upholds truth, respects the dignity of others, and avoids deception. Misinterpreting or distorting data could be seen as treating people (or their work) merely as a means to an end, rather than respecting their autonomy and their right to accurate and truthful representation. ant's ethical theory emphasizes respect for persons—each individual should be treated as an end in themselves, never merely as a means to an end. The duty to interpret properly aligns with this idea in that individuals have a right to accurate, honest, and clear interpretations of information that affects them. For example, in healthcare or legal contexts, individuals rely on others to interpret medical data, laws, or contracts. If these interpretations are done improperly or maliciously, individuals can be harmed, and their autonomy compromised. In Kantian ideology, interpreting data, laws, or communications incorrectly would be disrespectful because it fails to treat individuals as autonomous agents capable of making informed decisions. Misinterpreting data, especially in areas that directly affect people's lives (e.g., health care, criminal justice, or business), could lead to decisions that have negative consequences for individuals, violating the Kantian principle of respect for persons. Thus, a Kantian Deontologist would argue that we have an obligation to interpret data, facts, or communications correctly because this duty is a matter of universal moral law. If we fail in this duty, we would be contributing to a breakdown in the social fabric, undermining trust, and harming individuals' ability to act on truthful and reliable information. Kantian ethics also emphasizes moral accountability—we are responsible for our actions and their consequences, especially when our decisions affect others. Improper interpretation of data is a failure in this duty, as it involves misjudging or misrepresenting information in a way that can lead to harm or injustice. For instance, in a clinical setting, a misinterpretation of patient data could lead to incorrect diagnoses or treatments, directly affecting the patient's well-being. In summary, a Kantian Deontologist would defend the claim that we have a responsibility to proper interpretation of information as an obligation or duty because proper interpretation is not merely a best practice or an optional behavior; it is a moral obligation that we are duty-bound to uphold.