Hw 7

Dylan Doby

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 $\hat{\pi}$.

Student Answer

For a randomized response mechanism using a biased coin with probability of landing heads, the observed proportion of affirmative responses, $\hat{\pi}$, can be expressed as:

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

Rearranging this equation to isolate $\hat{P}$:

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

This is the generalized formula for $\hat{P}$. This formula accounts for any biased coin by adjusting the observed proportion of affirmative responses ($\hat{\pi}$) based on the probability of the coin landing heads (θ).

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

Student Answer

Special Case: $\theta = \frac{1}{2}$

To verify that the generalized formula reduces to the result from class when $\theta = \frac{1}{2}$, we substitute $\theta = \frac{1}{2}$ into the generalized formula:

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

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

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

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

Simplify the denominator and the term inside the parentheses:

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

$$\hat{P} = \frac{\hat{\pi} - 0.25}{0.5}$$

Divide both terms in the numerator by 0.5:

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

This is exactly the result derived in class for the fair coin case:

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

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

#chebychev function
chebychev <- function(x, y) {
  if (length(x) != length(y)) {
    stop("Vectors x and y must have the same length.")
  }
  # Calculate and return the maximum absolute difference
  max(abs(x - y))
}

#nearest_neighbors function
nearest_neighbors <- function(data, observation, k, distance_function) {
  # Ensure data is a matrix
  data <- as.matrix(data)

  # Calculate distances using the specified distance function
  distances <- apply(data, 1, function(row) distance_function(row, observation))

  # Sort distances and find the k-th smallest distance
  sorted_distances <- sort(distances)
```

```
  kth_distance <- sorted_distances[k]

  # Include all neighbors with distance <= k-th smallest distance
  nearest_indices <- which(distances <= kth_distance)

  # Return the indices and distances of the neighbors
  list(neighbor_list = nearest_indices, distances = distances[nearest_indices])
}

# Test Chebychev distance
x<- c(3,4,5)
y<-c(7,10,1)
chebychev(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)
#student input

# knn_classifier function
knn_classifier <- function(neighbors_data, label_column) {
  # Extract the class labels of the nearest neighbors
  class_labels <- neighbors_data[, label_column]

  # Find the mode (most frequent class label)
  mode_label <- names(which.max(table(class_labels)))

  return(mode_label)
}

#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
## 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')
```

```
## [1] "virginica"
```

```
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?

Student Answer

The output indicates that the predicted class for the observation is "virginica," and the true class label for this observation is also "virginica," confirming that the classification is correct. This demonstrates that the knn_classifier function successfully assigned the correct class label based on the nearest neighbors.

Although K=5 was specified, the output includes 7 observations due to ties in the Chebychev distance, where multiple observations share the same distance to the target observation as the K-th nearest neighbor. The Chebychev distance, which calculates the maximum absolute difference between corresponding dimensions of two vectors, inherently allows for ties when multiple points are equidistant from the target. In this implementation, all tied observations with distances equal to or less than the K-th smallest distance are included in the output. This means that no relevant neighbors are excluded arbitrarily, resulting in the 7 observations despite the specified k=5.

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

Sensitive healthcare data should remain strictly accessible to patients and their healthcare providers to preserve autonomy and ensure that patient well-being remains the focus. Allowing data transfer without

explicit patient consent, particularly in cases of company acquisition, undermines patient privacy as well as autonomy and jeopardizes the trust required for effective care.

From a deontological perspective, data transfer to a new company or to insurance providers would treat patients as mere means to corporate or actuarial ends rather than as individuals with inherent dignity. This is incompatible with Kantian Deontological ethics, which demands respect for persons and transparency in decisions that impact their lives. Furthermore, transferring data to insurance companies risks creating disparities in care. While insurance companies could leverage this data to refine actuarial models, they might also use it to deny coverage or increase premiums for vulnerable populations, exacerbating inequality.

Federated learning offers a viable solution by enabling the use of aggregate data for improving models without compromising individual privacy. This method aligns with Rawls's Veil of Ignorance, as it seeks to create fair outcomes without revealing protected characteristics. However, it is important to recognize that federated learning is not a perfect solution. Despite its design to keep sensitive data localized, it remains susceptible to risks such as data extraction attacks and model inversion techniques, which could reconstruct sensitive information from aggregated updates.

Therefore, while federated learning offers a promising approach to balancing innovation and privacy, its limitations highlight the need for stronger ethical safeguards and oversight. The risk of privacy breaches, even within aggregated frameworks, emphasizes the importance of keeping sensitive healthcare data under the direct control of patients and their healthcare providers. Any use of this data must uphold patient autonomy and fairness, with explicit consent as a fundamental requirement of ethical practice. By reinforcing protections against misuse and ensuring transparency, healthcare systems can build trust in technological advancements while safeguarding individual dignity and preventing the deepening of systemic inequalities.

I have described our responsibility to proper interpretation as an *obligation* or *duty*. How might a Kantian Deontologist defend such a claim?

Student Answer

A Kantian Deontologist would defend the claim that proper interpretation is a duty by appealing to the formulations of the categorical imperative, which demand universalizability and respect for moral agents. The obligation to properly interpret data ensures that researchers do not make claims that cannot be substantiated, as doing so would lead to outcomes that violate the principle of universalizability. If all researchers advanced unverified or improperly substantiated claims, public trust in the discipline would erode, rendering the entire pursuit of knowledge self-defeating. This aligns with Kant's argument that moral maxims must be applicable universally without contradiction.

Furthermore, Kantian ethics emphasizes treating others as ends in themselves, not merely as means to an end. Advancing claims without proper interpretation would exploit the trust of both fellow researchers and the public for personal or professional gain, thereby violating their dignity as rational moral agents. Researchers who misuse statistical interpretations for sensationalism or convenience fail to respect their audience's capacity for reasoned judgment and informed decision-making.

The duty to proper interpretation also stems from the potential consequences of mishandling data, such as perpetuating misinformation or compromising subsequent research based on flawed conclusions. By adhering to rigorous interpretive standards, researchers uphold the integrity of their discipline and ensure that their work respects the autonomy and intellectual dignity of others. This duty is not merely a matter of professional ethics but a moral imperative rooted in Kantian principles, signifying the profound responsibility of those who generate and communicate knowledge.