STOR390 HW7

Rebekah Kirkman

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

$$\hat{P} = \frac{\theta}{2} + \frac{1}{2}$$

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

To show that the expression for $\hat{P}$ reduces to our result from class when $\theta = \frac{1}{2}$, we substitute $\theta = \frac{1}{2}$ into the generalized expression:

$$\begin{aligned} \hat{P} &= \frac{\theta}{2} + \frac{1}{2} \\ &= \frac{\frac{1}{2}}{2} + \frac{1}{2} \\ &= \frac{1}{4} + \frac{1}{2} \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1 \end{aligned}$$

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

So, when $\theta = \frac{1}{2}$, the expression for $\hat{P}$ simplifies to $\hat{P} = 1$.

This confirms that our result from class, $\hat{P} = 2\pi - \frac{1}{2}$, is a special case of the generalized expression when $\theta = \frac{1}{2}$.

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.

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

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
cheby <- function(x, y) {
  if (length(x) != length(y)) {
    stop("Vectors must be of equal length")
  }
  absolute_diff <- abs(x - y)
  max_absolute_diff <- max(absolute_diff)
  return(max_absolute_diff)
}

#nearest_neighbors function
nearest_neighbors <- function(data, query_point, k, distance_function) {
  distances <- apply(data, 1, function(x) distance_function(x, query_point))
  nearest_indices <- order(distances)[1:k]
  return(data[nearest_indices, ])
}

x<- c(3,4,5)
```

```r
y<-c(7,10,1)
cheby(x,y)
```

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(nearest_neighbors, class_column) {
  neighbor_labels <- nearest_neighbors[, class_column]
  mode_label <- names(sort(table(neighbor_labels), decreasing = TRUE))[1]
  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])
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?

The output dataframe contains 7 observations instead of the specified k=5 because multiple observations have the same distance from the query point, and in such cases, all observations with the same minimum distance are included 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.

Access to sensitive healthcare data should be restricted to those with a legitimate need for it, such as healthcare providers directly involved in patient care and any transfer or sharing of data should follow strict consent processes and privacy protections. From a deontological perspective, individuals have inherent rights and dignity that must be respected, regardless of the potential benefits to society. In the context of healthcare data, patient autonomy and privacy are important moral considerations because patients have a fundamental right to control their personal health information and make informed decisions about its use. Healthcare providers have a duty as professionals to safeguard patient confidentiality and only access data that is necessary for providing care. When it comes to third-party entities such as software management companies or insurance providers, access to sensitive health data must be governed by strict ethical principles and legal regulations. Health data may be useful in improving healthcare or informing insurance practices, but ultimately, the rights and interests of patients must be at the forefront of all ethical considerations around using health data.