Homework 8

MTH 3270 Data Science

Due Mon., Apr. 18

Read These Chapters of the Book	**Then Do These Exercises**
11	Problem 6* (parts ***a*** and ***c*** only, and just do **neural network** (**Ch 11**)
12	6** (part ***a*** only) (**Ch 12**), Problem 1 (**below**)

* For **Problem 6** (**Ch 11**):

- The `NHANES` data set is in the `"NHANES"` package. The help page has a description of the data set:

```
library(NHANES)
? NHANES
```

- There are *many* `NA`s in the data set. In fact, *every row* has at least one `NA`:

```
any(complete.cases(NHANES))
```

  One way to deal with the `NA`s is to *first* use `select()` (from `"dplyr"`) to create a new data frame containing only the variables (columns) from `NHANES` that you want to use in your classification models, *then* use `na.omit()` (or `complete.cases()`) to create a new version of that data frame which contains only the observations (rows) that don't have any `NA`s.

- **Don't** use any of the **categorical** variables as explanatory variables (X's) in the classification model. To see which variables are **numeric** (or **integer**) and which are **categorical** (**factors**), type:

```
str(NHANES)
```

- It's *possible* that your **neural network** might end up predicting everyone to be in one class. If this happens, try changing the value of the **complexity parameter** (or **tuning parameter**), i.e. the number of **hidden units** k – see Exercise 10 in Class Notes 6.

** For **Problem 6** (**Ch 12**) , you can copy and paste the R code below (instead of typing it yourself):

```
# install.packages("mdsr")
library(mdsr)

# install.packages("Lahman")
library(Lahman)

hof <- Batting %>% group_by(playerID) %>%
  inner_join(HallOfFame, by = c("playerID" = "playerID")) %>%
  filter(inducted == "Y" & votedBy == "BBWAA") %>%
  summarize(tH = sum(H), tHR = sum(HR), tRBI = sum(RBI), tSB = sum(SB)) %>%
  filter(tH > 1000)
```

1 Repeat **Problem 6** (part ***a*** only) from **Ch 12**, but now use `hclust()` (instead of `kmeans()`) to perform **hierarchical clustering**.