CSC 587 HW 3

Daniel R. Getty

2024-02-20

Set r Environment

```
knitr::opts_chunk$set(echo = TRUE, message = TRUE)
# directory
dir <- 'G:\\My Drive\\H Drive\\Course Work\\CERG-Data Science\\CSC_587_Advanced_Data_Mining\\HW\\HW3_No
# Set the working directory.
setwd(dir)
# Print the working directory.
getwd()
```

```
## [1] "G:/My Drive/H Drive/Course Work/CERG-Data Science/CSC_587_Advanced_Data_Mining/HW/HW3_Normalize
```

```
# load ggplot2 package
library(ggplot2)
# load ggplot2 package
library(ggplot2)
# load dplyr package
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# load tidyr package
library(tidyr)
```

Set py Environment

```
import os
import pandas as pd
import numpy as np
import scipy.stats as stats
import math
```

Homework 3

2. write a function in your preferred language which can take a data vector and do min-max normalization by transforming data onto a desired range. For example, it should be able get the age data above and map it between any two numbers. where a is an one-dimensional array)

```
# Using Python
# Min-Max Normalization function
def min_max_norm(data, new_min, new_max):
    old_min, old_max = min(data), max(data)
    xscaled = [new_min + (x - old_min) * (new_max - new_min) / (old_max - old_min) for x in data]
    return xscaled

# Create Data Vector of ages
age = [13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 4

# Perform Min-Max Normalization on age
new_min = 0
new_max = 1

# Call the min_max_norm function
age_norm = min_max_norm(age, new_min, new_max)

# zip the age and age_norm
age_age_norm = list(zip(age, age_norm))

# Print the normalized age
print("(Age, Age_Norm)", age_age_norm)
```

```
## (Age, Age_Norm) [(13, 0.0), (15, 0.03508771929824561), (16, 0.05263157894736842), (16, 0.052631578947
```