Transform education to dummy variable remove education original and change it to dummy variable for predictor variable

13 variables in the end result

Normalize line 108 copy of original and normalized data ---- find his example from somewhere need to keep original data to refer to see line 113 --- train.norm.df find his examples

Create the model – pre-process then define the model and the test data must be normalized normalize training, validation, and test

Personal loan and education are the only ones to convert to factors

Education <- dummyVars(~Education,ubfd)  
EduDV <- predict(Education,ubfd)

Grouping variable is the income

Question was – can you tell us what steps to take in order? 🡨 there was complete silence at this point…

1. Read

 Read and clean input data; this includes converting variables to factor, and then to dummy variables.

 Split the data into training and validation

 Normalize the data

 Answer Question 1 by creating the test data, apply the normalization on the test data

 You will adopt the following process in general:

 Split data; Normalize using training data; Find best K; Combine Training and Validation; Renormalize and then apply to the normalized test data



Create 2 confusion matrix – one for 60/40 split and one for 50/30/20

After question 1 ---- K is equal to 1

Justify your answer – what does that mean?

Personal loan is your y or why? Response is personal income the others are the predictors

---

title: "Assignment2"

output: pdf\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the \*\*Knit\*\* button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

install.packages("class")

```{r}

Myfile <- read.csv("UniversalBank.csv")

```

```{r}

library(dummies)

library(ISLR)

Myfile$Education = as.factor(Myfile$Education)

dummy = dummy.data.frame(select(Myfile,-c(ZIP.Code,ID)))

dummy$CreditCard = as.factor(Myfile$CreditCard)

dummy$Personal.Loan = as.factor(Myfile$Personal.Loan)

dummy

```

```{r}

set.seed(123)

train.index= sample(row.names(dummy), 0.6\*(dummy))

valid.index=setdiff(row.names(dummy),train.index)

tdf = dummy[train.index]

vdf = dummy[valid.index]

```