Class_2_Homework_Hints

```
setwd("C:/Users/ellen/Documents/UH/Spring 2020/DA2/Section 1/MIdTerm")

rmse <- function(error)
{
  sqrt(mean(error^2))
}


set.seed(223)

ProductSales = read_csv("ProductSalesv2.csv")

ProductSales$WkBeg = mdy(ProductSales$WkBeg)

ProductSales = pivot_longer(ProductSales, 3:5, names_to = "Product", values_to = "Sales")

p <- ggplot(data = ProductSales, aes(WkBeg, Sales, color = Product)) + geom_point() +
  theme(panel.background = element_rect(fill = "white"))
p
```


```
Train = ProductSales %>% filter(WkBeg < "2015-01-01")
Test  =  ProductSales %>%  filter(WkBeg >= "2015-01-01")
```

```r
# lm polynomial  ----------------- #

TestMod2 = lm(Sales ~ Product + Wk + I(Wk^2), data = Train)

rmse2 = rmse( Test$Sales - predict(TestMod2, Test))
rmse2
```

```
## [1] 13.88266
```

```r
mXPoly = model.matrix(Sales ~ Product + Wk + I(Wk^2), data = Train)
vY = as.numeric(Train$Sales)
vBetaPoly <- solve(t(mXPoly)%*%mXPoly, t(mXPoly)%*%vY) # solve using normal equations
yPoly = t(as.numeric(vBetaPoly)%*%t(mXPoly))

# -- test

mXPolyTest = model.matrix(Sales ~ Product + Wk + I(Wk^2), data = Test)

rmse4 = rmse( Test$Sales - (t(as.numeric(vBetaPoly)%*%t(mXPolyTest))))

mXPolyRMSE = data.frame(Method = "mXPoly", RMSE = rmse4, stringsAsFactors = F)


# Regularization

n = ncol(mXPoly)

d = diag(1,n,n)
d[1,1] = 0

#th = array(0,c(n,length(lambda)))

vBetaReg1 = as.numeric(solve(t(mXPoly) %*% mXPoly + (40 * d)) %*% (t(mXPoly) %*% vY))
vBetaReg2 = as.numeric(solve(t(mXPoly) %*% mXPoly + (60 * d)) %*% (t(mXPoly) %*% vY))
vBetaReg3 = as.numeric(solve(t(mXPoly) %*% mXPoly + (80 * d)) %*% (t(mXPoly) %*% vY))

rmseReg1 = rmse( Test$Sales - (t(as.numeric(vBetaReg1)%*%t(mXPolyTest))))
rmseReg2 = rmse( Test$Sales - (t(as.numeric(vBetaReg2)%*%t(mXPolyTest))))
rmseReg3 = rmse( Test$Sales - (t(as.numeric(vBetaReg3)%*%t(mXPolyTest))))

mXPolyRMSE = data.frame(Method =  c(
  "mXReg1",
  "mXReg2",
  "mXReg3"
),

RegPenalty = c(40, 60, 80),

RMSE =  c(
  rmseReg1,
  rmseReg2,
  rmseReg3), stringsAsFactors = F)
```

```
knitr::kable(mXPolyRMSE) %>%
  kable_styling(full_width = F, bootstrap_options = "striped", font_size = 9)
```

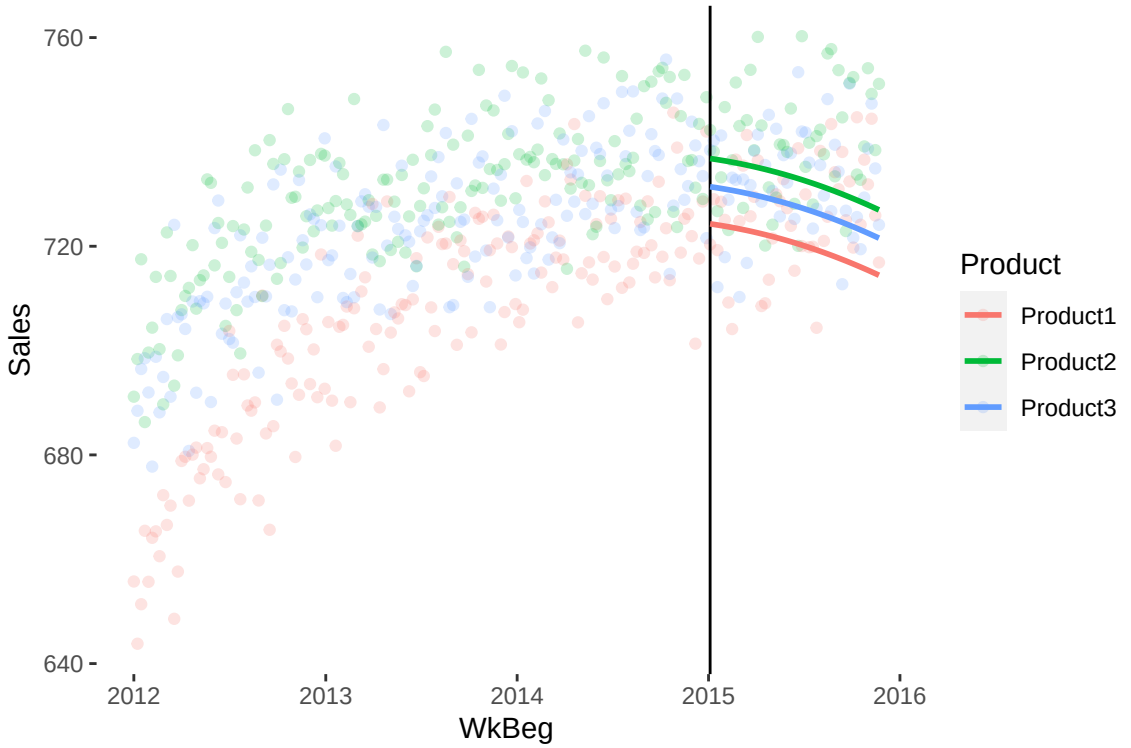
Method	RegPenalty	RMSE
mXReg1	40	12.87197
mXReg2	60	12.84000
mXReg3	80	12.86688

So the penalty of 60 got an rmse under 12.85

Solving:

```
Test$Pred =  t(as.numeric(vBetaReg2)%*%t(mXPolyTest))

p <- ggplot(data = ProductSales, aes(WkBeg, Sales, color = Product)) + geom_point(alpha = .2) +
  theme(panel.background = element_rect(fill = "white"))
p = p + geom_vline(xintercept = as.Date("2015-01-04"))
p <- p + geom_line(data = Test, aes(x = WkBeg, y = Pred, color = Product), size = 1)
p
```



Now, to find the peak of the PLM

first, get vBetafromBestModel:

```
vBetaReg2
```

```
## [1] 679.184266879  12.520746037   7.112782350   0.672441885  -0.002449081
```

So, that means the linear equation is $679.184266879 + 12.520746037 * Product2 + 7.112782350 * Product3 + 0.672441885 * Wk - 0.002449081 * Wk^2$

So, the derivative is:

$(2 * .002449081)Wk = 0.672441885$

So, Wk =

$Wk = 0.672441885/(2 * .002449081)$ $Wk = 137$