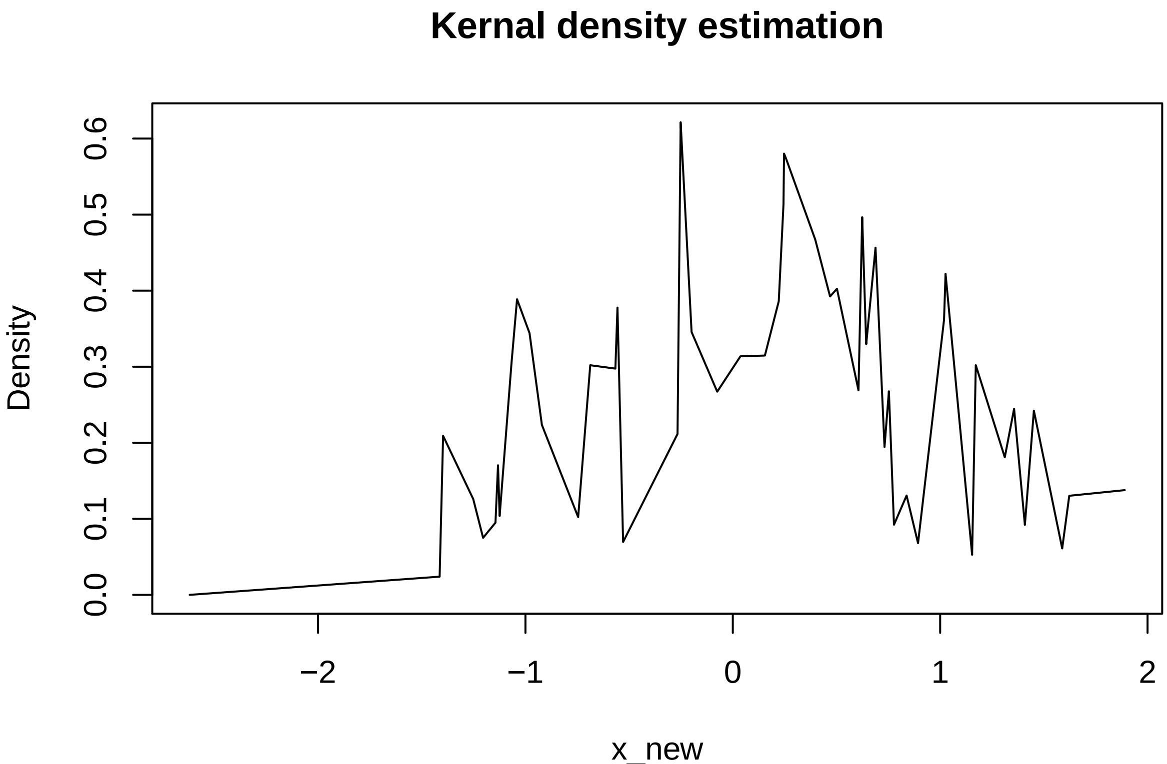
Homework 3

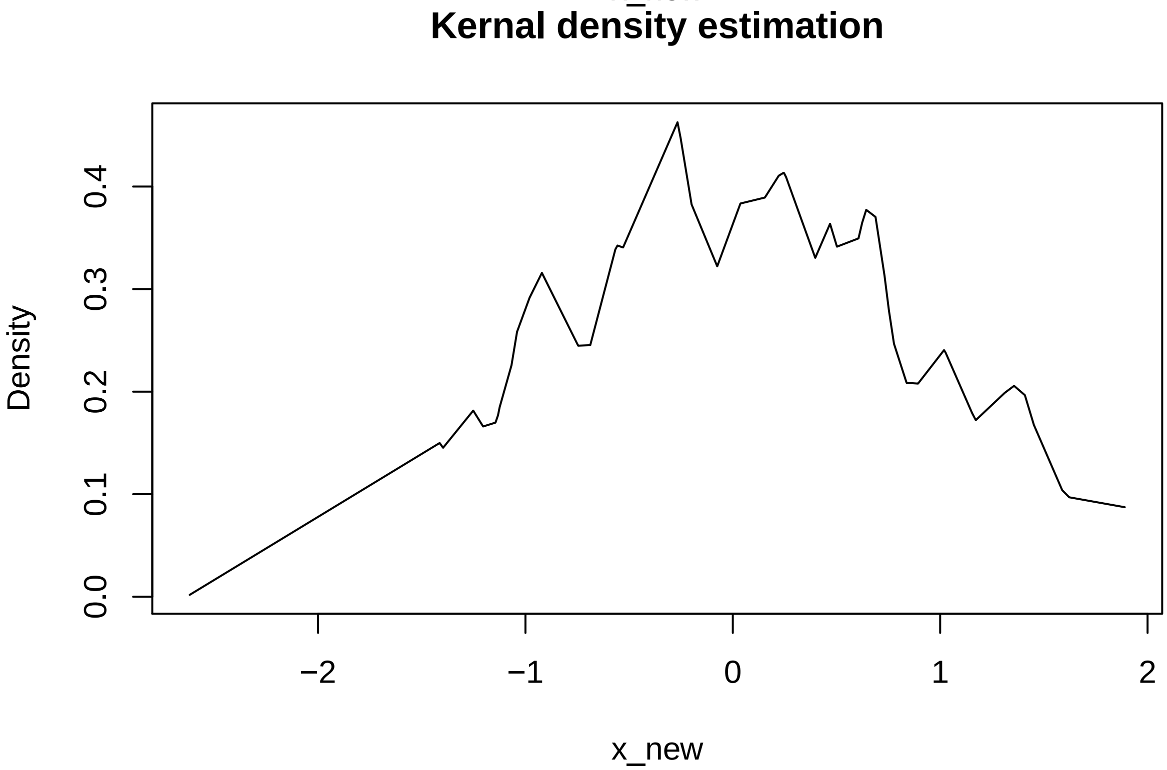
This homework is due by the end of the day on November 5th 2018. Solutions should appear as a vignette in your package called "homework-3". If you do not want to write your solution for questions 2 - 4 in LaTeX, you may provide a solution using paper and pencil and include an image in your write-up.

1. CASL page 117, question 7.
2. CASL page 200, question 3
3. CASL page 200, question 4
4. CASL page 200, question 5
5. CASL page 200, question 6

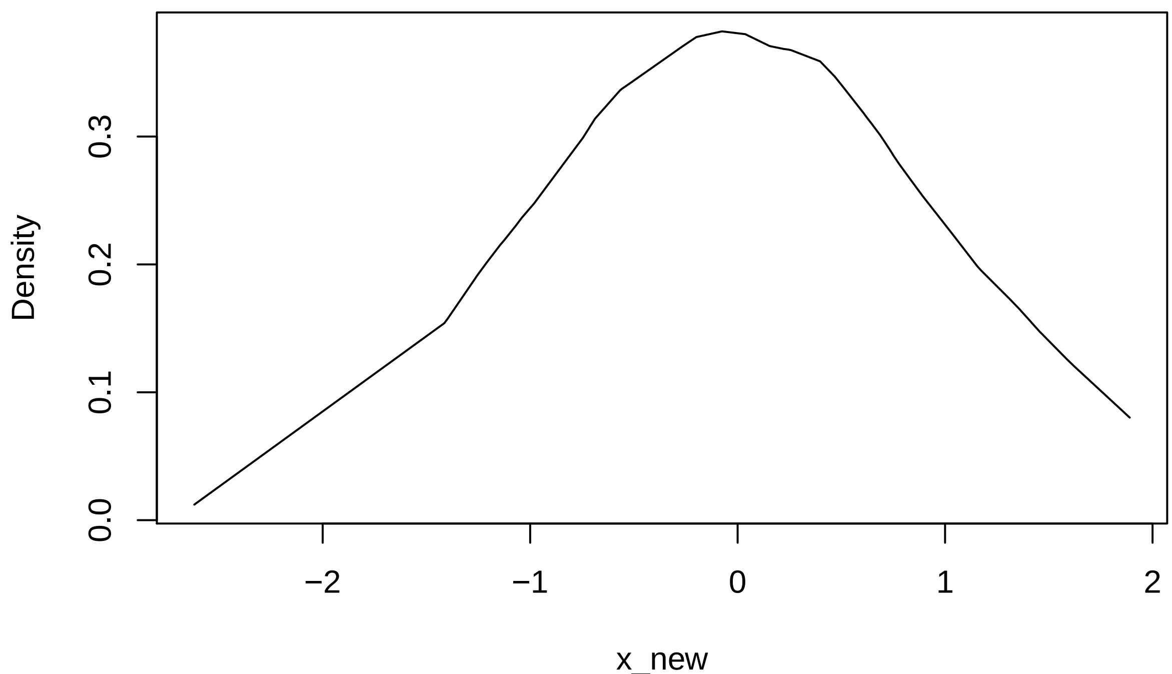
1.

```
set.seed(6909)
#write out the Epanechnikov kernel function
kernel_epan <- function(x) {
ran <- as.numeric(abs(x) <= 1)
val <- (3/4) * ( 1 - x^2 ) * ran
return(val)
}

#write out the Epanechnikov density function
i = 1
h = 1
kern_density <- function(x, h, x_new){
  dst_est <- numeric()
  for (i in 1:length(x_new)){
    dst_est[i] <- mean(kernel_epan((x_new[i]-x)/h))/h
  }
  dst_est
}

#visualize different bandwidths
x <- rnorm(1000, 0, 1)
x_new <- sort(rnorm(50, 0, 1))
h = c(0.01, 0.1, 0.5, 1, 2, 3)
for (i in h){
  plot(x_new, kern_density(x,i,x_new), ylab = "Density", main = "Kernal density estimation", type = "l")
}
```

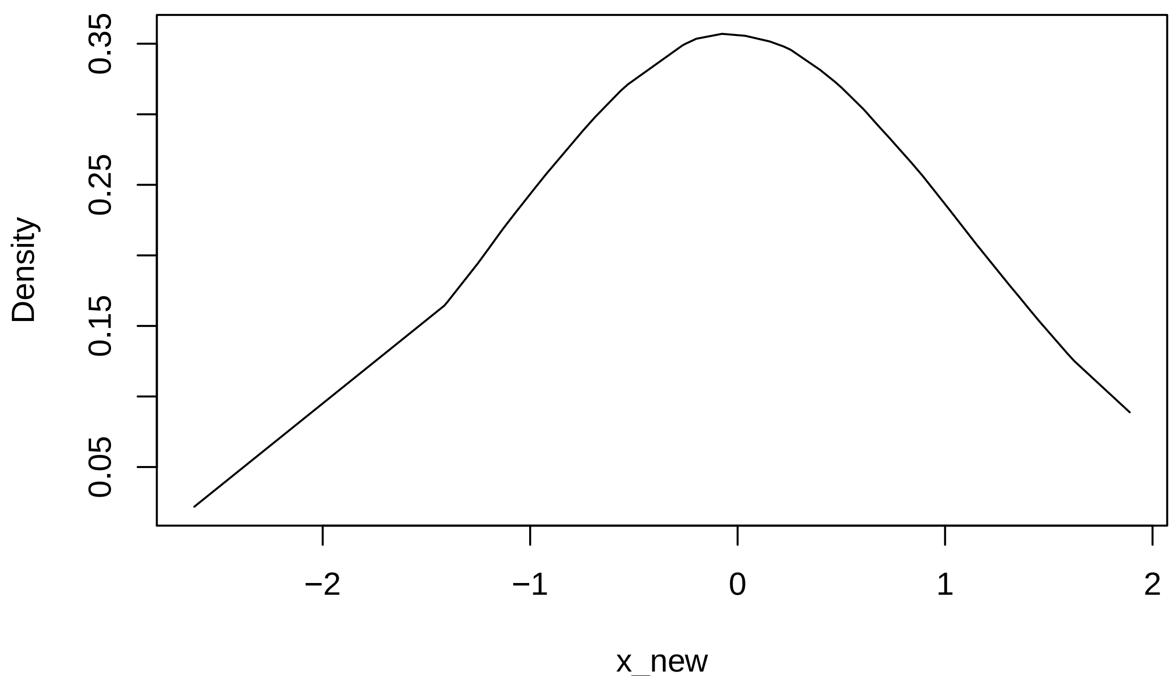

Kernal density estimation
Density
x_new
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−2
−1
0
1
2

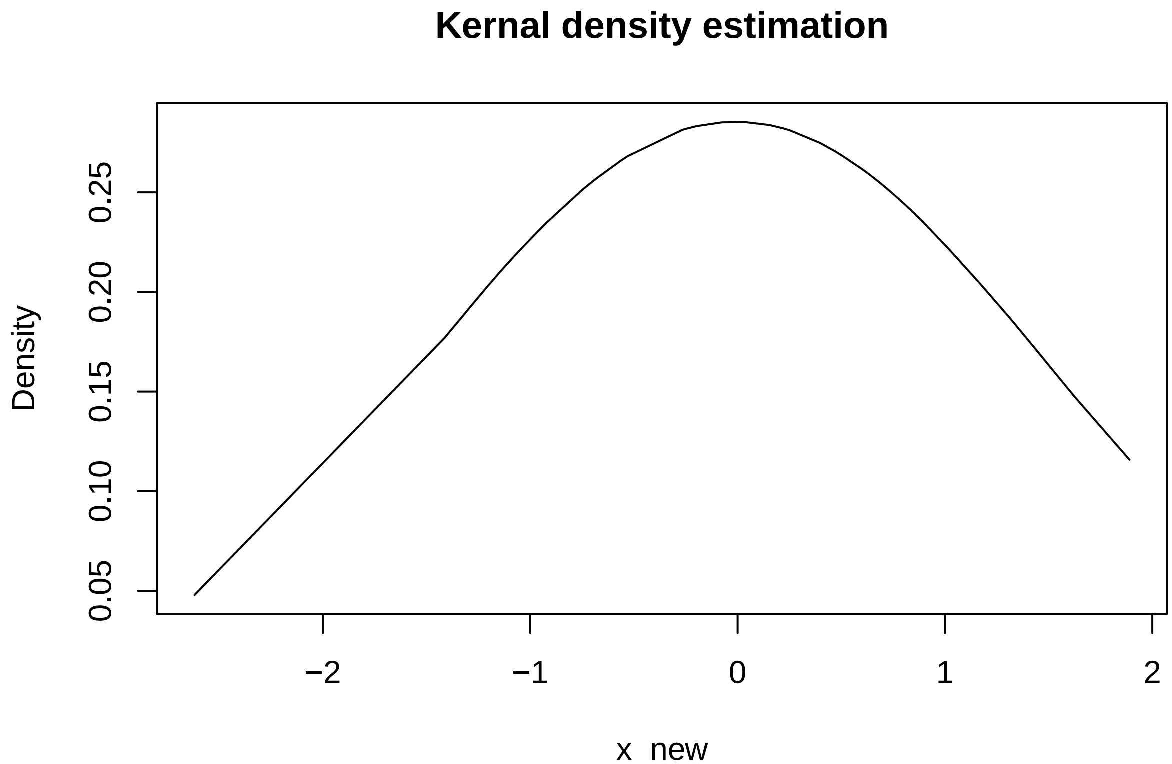

Kernal density estimation
Density
x_new
0.0
0.1
0.2
0.3
0.4
−2
−1
0
1
2

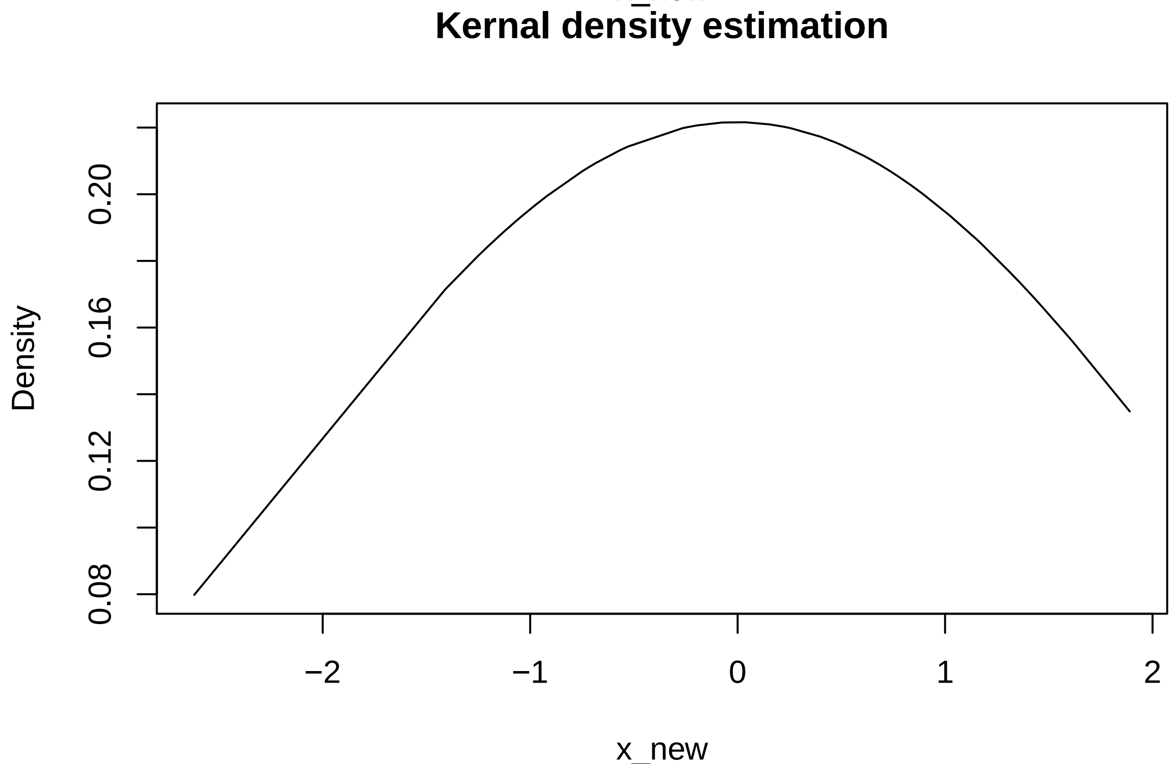
Kernal density estimation



Kernal density estimation







From the plots above, as bandwidth increases, the estimations become smoother and closer to normal distribtion. And when bandwidth increases to a certain value, the estimation remains similar no matter how bandwidth increases.

2.

By definition of convex function, for $t \in [0,1]$

$$\begin{aligned} f(tx+(1-t)y) &\leq tf(x)+(1-t)f(y) \\ g(tx+(1-t)y) &\leq tg(x)+(1-t)g(y) \end{aligned}$$

Then

$$\begin{aligned} h(tx+(1-t)y) &= f(tx+(1-t)y)+g(tx+(1-t)y) \\ &\leq tf(x)+(1-t)f(y)+tg(x)+(1-t)g(y) \\ &= t(f(x)+g(x))+(1-t)(f(y)+g(y)) \\ &= th(x)+(1-t)h(y) \end{aligned}$$

Therefore, $h=f+g$ is convex.

3.

we want to prove that $f(x)=|x|$ is convex. For $t \in [0,1]$

$$\begin{aligned} f(x) &= |x| \\ f(tx+(1-t)y) &= |tx+(1-t)y| \\ &\leq |tx|+|(1-t)y| \\ &= f(tx)+f((1-t)y) \\ &= tf(x)+(1-t)f(y) \end{aligned}$$

Thus, $f(x)=|x|$ is convex.

the other question, $\mathcal{L}_1$ norm:

$$|x|_1 = \sum_{i=1}^{n} |x_i|$$

By the last question, it's also convex.

4.

l_2 *norm*

$$\begin{aligned} f(x) &= x^2 \\ f(tx+(1-t)y) &= t^2x^2+(1-t)^2y^2+2t(1-t)xy \\ tf(x)+(1-t)f(y) &= tx^2+(1-t)y \\ f(tx+(1-t)y)-[tf(x)+(1-t)f(y)] &= t(t-1)x^2+t(t-1)y^2+2t(1-t)xy \\ &= t(1-t)(-x^2-y^2+2xy) \\ &= t(t-1)(x-y)^2 \\ &\leq 0 \ \ \textit{since } 0 \leq t \leq 1 \\ f(tx+(1-t)y) &\leq tf(x)+(1-t)f(y) \end{aligned}$$

Therefore, $f(x)=x^2$ is convex. Objective function of elastic net is

$$\frac{1}{2n}||y-Xb||_2^2+\lambda[(1-\alpha)\frac{1}{2}||b||_2^2+\alpha||b||_1]$$

It's a sum of l_2 *norm* and l_1 *norm*, from the previous two exercises,it's convex

5.

According to textbook page 189:

```r
library(glmnet)
```

```
## Loading required package: Matrix
## Loading required package: foreach
## Loaded glmnet 2.0-16
```

```r
# KKT check function
#Return:
#A logical vector indicating where the KKT conditions have been violated by the variables that are curr
check_kkt <- function(y, X, b, lambda) {
  resids <- y - X %*% b
  s <- apply(X, 2, function(xj) crossprod(xj, resids)) / lambda / nrow(X)
  (b == 0) & (abs(s) >= 1)
}

##use iris as dataset

x <- scale(model.matrix(Sepal.Length ~. -1, iris))
y <- iris[,1]


#implement lasso regression

#Check which variables violate KKT condition
#Return: logical vector indicating where the KKT conditions have been violated
lasso_reg_with_screening <- function(x, y){
  m1 <- cv.glmnet(x,y,alpha=1)
  #use 1se as the criteria to choose lambda
  lambda <- m1$lambda.1se
  b <- m1$glmnet.fit$beta[, m1$lambda == lambda]
  check_kkt(y, x, b, lambda)
}
lasso_reg_with_screening(x,y)
```

```
##       Sepal.Width      Petal.Length       Petal.Width     Speciessetosa
##             FALSE             FALSE             FALSE             FALSE
## Speciesversicolor  Speciesvirginica
##             FALSE             FALSE
```

We can observe thatall coefficients get a false result for KKT violation,which indicates that no KKT violation