1. a) $X \sim N(0, \sigma^2)$

$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$

$Y = |X| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases} \Rightarrow \begin{matrix} x = y & \text{if } x \geq 0 \\ x = -y & \text{if } x < 0 \end{matrix}$

$f_Y(y) = f_X(y) \cdot \left|\frac{d}{dy} y\right|$

$= \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$ if $x \geq 0$ $\quad y \geq 0$

$f_Y(y) = f_X(-y) \cdot \left|\frac{d}{dy} -y\right|$

$= \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$ if $x < 0$ $\quad y \geq 0$

Thus $f_Y(y) = \frac{2}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$ for $y \geq 0$

b) $E(y) = \int_0^\infty y f_Y(y)\, dy = \int_0^\infty y \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma} e^{-y^2/2\sigma^2} dy$

let $-y^2/2\sigma^2 = t$
$y\,dy = \sigma^2 dt$

$E(y) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \sigma^2 \int_0^\infty e^{-t} dt$

$= \left[\sqrt{\frac{2}{\pi}} \sigma (-e^{-t})\right]_0^\infty$

$\boxed{= \sqrt{\frac{2}{\pi}} \sigma}$

$E(y^2) = \int_0^\infty y^2 \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} e^{-y^2/2\sigma^2} dy$

$= \sqrt{\frac{2}{\pi}} \sigma \int_0^\infty \sqrt{2\sigma^2 t}\, e^{-t} dt$

$= \frac{2}{\sqrt{\pi}} \sigma^2 \int_0^\infty t^{\frac{1}{2}} e^{-t} dt \Rightarrow$ Gamma dist

$= \frac{2}{\sqrt{\pi}} \sigma^2 \Gamma(\frac{3}{2})$

$= \frac{1}{\sqrt{\pi}} \sigma^2 \Gamma(\frac{1}{2})$

$= \sigma^2$

$Var(y) = E(y^2) - E(y)^2$
$= (1 - \frac{2}{\pi}) \sigma^2$

c) $L(\sigma^2 | \underset{\sim}{y}) = \prod_{i=1}^{N} f_Y(y_i) = \left(\frac{2}{\pi}\right)^{\frac{N}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}} e^{-\frac{\sum y_i^2}{2\sigma^2}}$

$\log L(\sigma^2 | \underset{\sim}{y}) = \frac{N}{2} \log \frac{2}{\pi} - \frac{N}{2} \log \sigma^2 - \frac{\sum y_i^2}{2\sigma^2}$

$$\frac{d \log L(\sigma^2 | \underset{\sim}{y})}{d\sigma^2} = -\frac{N}{2\sigma^2} + \frac{\sum y_i^2}{2(\sigma^2)^2} = 0$$

$$\frac{\sum y_i^2}{\cancel{2}(\sigma^2)^{\cancel{2}}} = \cancel{\frac{N}{2\sigma^2}}$$

$$\text{MLE}: \hat{\sigma}^2 = \frac{\sum y_i^2}{N}$$

2. a) Since $P(A, B) = P(A) P(B)$, with unobserved check point C,

Then $P(A | B) = P(A)$, $P(B | A) = P(B)$

Then $P(A | B, \phi) = P(A | \phi)$

Thus $A \perp\!\!\!\perp B \mid \phi$

b) $P(A, B, C, D) = P(A) P(B) P(C | A, B) P(D | C)$

$$P(A, B, C | D) = \frac{P(A) P(B) P(C | A, B) P(D | C)}{P(D)}$$

$$P(A, B | D) = \frac{P(A) P(B) \cdot \sum_C P(C | A, B) P(D | C)}{P(D)}$$

Then $P(A | B, D) = \dfrac{P(A) \cdot \sum P(C | A, B) P(D | C)}{P(D)} \neq P(A | D)$

Thus $A \not\perp\!\!\!\perp B \mid D$

c) $P(C=1 \mid d=2, a=0, b=1) = \dfrac{P(a,b,c,d)}{P(a,b,C=0,d)+P(a,b,C=1,d)}$

$$= \frac{P(a=0)P(b=1)P(c=1 \mid a=0,b=1)P(d=2 \mid c=1)}{P(a=0)P(b=1)P(c=1 \mid a=0,b=1)P(d=2 \mid c=1)+P(a=0)P(b=1)P(c=0 \mid a,b)P(d=2 \mid c=0)}$$

$$= \frac{0.5\times0.5\times0.5\times0.00443}{0.5\times0.5\times0.5\times(0.00443+0.05399)}$$

$$= \frac{0.00443}{0.00443+0.05399}$$

$$= 0.0758$$

d) $P(d=2) = \sum_{a,b,c} P(a,b,c,d)$

$$= 0.5\times0.5\times\left[(0.1+0.1+0.3+0.5)\times0.05399 + (0.9+0.9+0.7+0.5)\times0.00443\right]$$

$$= 0.01682$$

$P(C=1) = \sum_{a,b} P(a,b,c) = 0.5\times0.5\times(0.9+0.9+0.7+0.5)$

$$= 0.75$$

$$P(C=1 \mid d=2) = \frac{P(d=2 \mid c=1)\,P(c=1)}{P(d=2)}$$

$$= \frac{0.00443\times0.75}{0.01682} = 0.0197$$

3. a)

$$\|t - Xw\|^2$$

$$L(\lambda_1, \lambda_2, w) = (t - Xw)^T (t - Xw) + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2$$

$$P(w) = \text{Lasso}(w) = \|t - X\hat{w}\|^2 + \lambda_1 \|\hat{w}\|_1$$

$$P_2(w) = \text{Ridge}(w) = \|t - Xw\|^2 + \lambda_2 \|w\|_2^2$$

$$P_3(x) = \text{Lineer} = \|t - Xw\|^2$$

b) ① let $l = 0$, $w_1^{(0)} = 0.2$, $w_2^{(0)} = 1$

$$x^{(l)} = \arg\min \|t - Xw^{(l)}\|_{l_1}$$

③ $$w_1^{(l+1)} = \arg\min \left[\|t - Xw^{(l)}\|_2^2 + \lambda_1 \|w^{(l)}\|_1 \right]$$

$$w_2^{(l+1)} = \arg\min \left[\|t - Xw^{(l)}\|_2^2 + \lambda_2 \|w^{l}\|_2^2 \right]$$

4. a) $p(x) = \sum_{k=1}^{K} \pi_k \, p(x|k)$

$$p(x_b, x_a) = \sum_{k=1}^{K} \pi_k \, p(x_b, x_a | k)$$

$$= \sum_{k=1}^{K} \pi_k \, p(x_b | x_a, k) \, p(x_a | k)$$

$$p(x_a) = \int_{x_b} \sum_{k} \pi_k \, p(x_b, x_a | k) \, dx_b$$

$$= \sum_{k} \pi_k \int_{x_b} p(x_b, x_a | k) \, dx_b$$

$$= \sum_{k} \pi_k \, p(x_a | k)$$

$$p(x_b | x_a) = \frac{p(x_b, x_a)}{p(x_a)}$$

$$= \frac{\sum_{k=1}^{K} \pi_k \, p(x_b, x_a | k)}{\sum_{k} \pi_k \, p(x_a | k)}$$

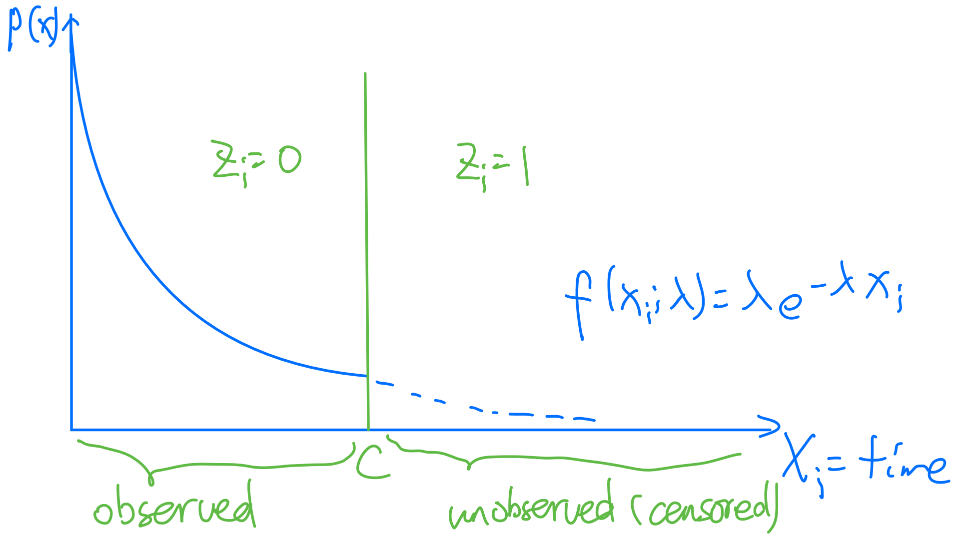
$$= \frac{\sum_{k=1}^{K} \pi_k \, p(x_b | x_a, k) \, p(x_a | k)}{\sum_{k=1}^{K} \pi_k \, p(x_a | k)}$$

$$= \sum_{k=1}^{K} \frac{\pi_k \, p(x_a | k)}{\sum_{k=1}^{K} \pi_k \, p(x_a | k)} \, p(x_b | x_a, k)$$

b) mixing coef: $\dfrac{\pi_k \, p(x_a | k)}{\sum_{k=1}^{K} \pi_k \, p(x_a | k)}$

component densities: $p(x_b | x_a, k)$

5. a) Since $x_1 \dots x_n \sim \exp(\lambda)$



b) $$P(X=x \mid X \geq c) = \frac{P(X=x, X>c)}{P(X \geq c)}$$

$$= \frac{f(x,\lambda)}{1 - F_X(x \geq c)}$$

$$= \frac{\lambda e^{-\lambda x}}{1-(1-e^{-\lambda c})}$$

$$= \lambda e^{-\lambda(x-c)}$$

C) $L(x_i, y_i, z_i, i=1..n : \lambda) = \prod_{i=1}^{n} p(x_i)^{1-z_i} p(y_i)^{z_i}$

$p(Y=y) = \lambda e^{-\lambda(y-c)}$

$$E[\log L(\cdot)] = E\left[\sum_{i=1}^{n} (1-z_i)\log p(x_i) + z_i \log p(y_i)\right]$$

$$= \sum_{i=1}^{n} (1-z_i)(\log\lambda - \lambda x_i) + z_i\, E(p(y=y_i))\, ((\log\lambda - \lambda y_i)$$

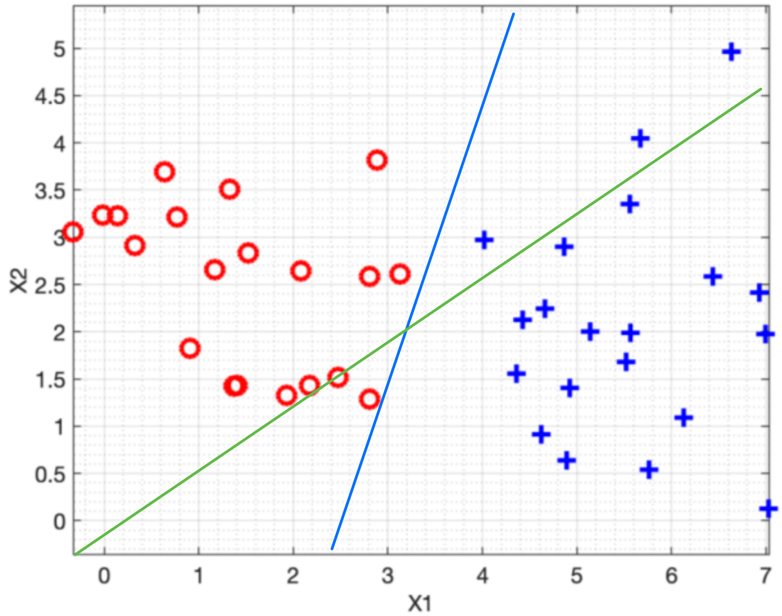
$$= \sum_{i=1}^{n} ((1-z_i)(\log\lambda - \lambda x_i) + z_i \log\lambda - \lambda(c + \frac{1}{\lambda^{old}})$$

$$Q(\lambda, \lambda^{old}) = n\log\lambda - \lambda \sum ((1-z_i)x_i + z_i(c + \frac{1}{\lambda^{old}})$$

$$\frac{\partial Q}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} ((1-z_i)x_i + z_i(c + \frac{1}{\lambda^{old}}) = 0$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} (1-z_i)x_i + z_i(c + \frac{1}{\lambda^{old}})}$$

6.



a) $x_2 = -\frac{x_1}{x_2} \cdot x_1 - \frac{1}{x_2}$

$x_2 = -\frac{x_1^2}{x_2} - \frac{1}{x_2}$

$x_1^2 + x_2^2 = 1$

The decision boundary would be a line, start at x_1-axis, w would be about 0.5 for each.

b) Since there is no constant term in this model. The decision boundary would become a line start from (0, 0) and w may not balance, may be about (0.2, 0.8) - (0.3 - 0.7)

c) $p(D|M_1) = \int p(D|w_1, M_1)\, p(w_1|M_1)\, dw_2$

$p(D|M_2) = \int p(D|w_1, M_1)\, p(w_1|M_1)\, dw_2$

midterm

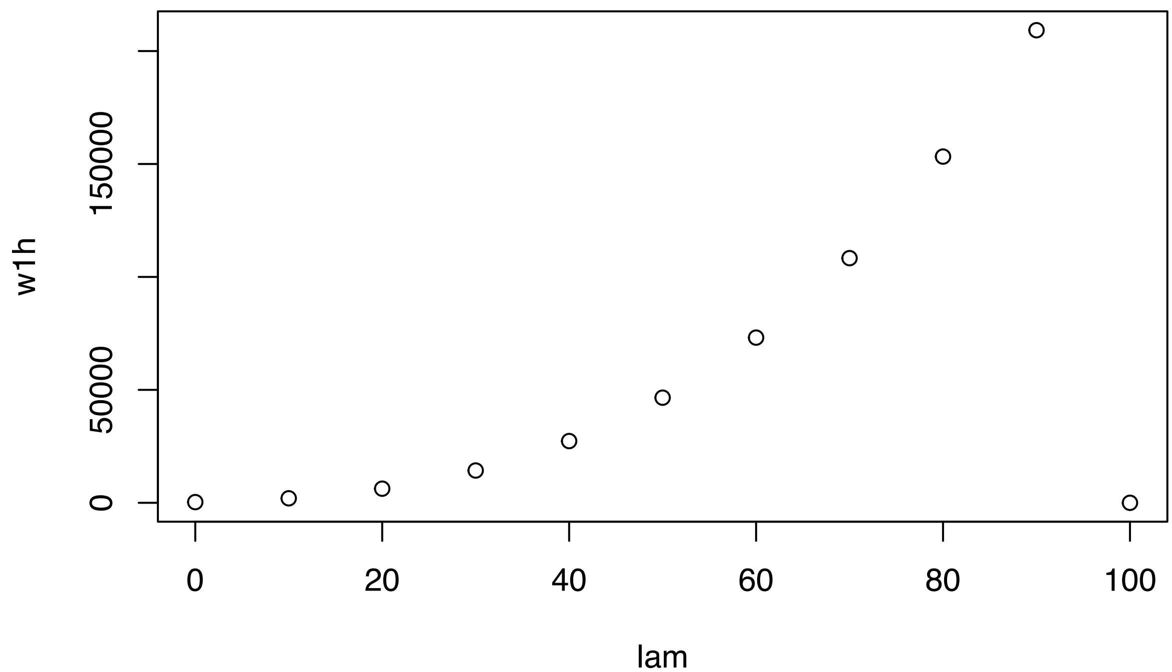
Enbo Tian

2022/3/4

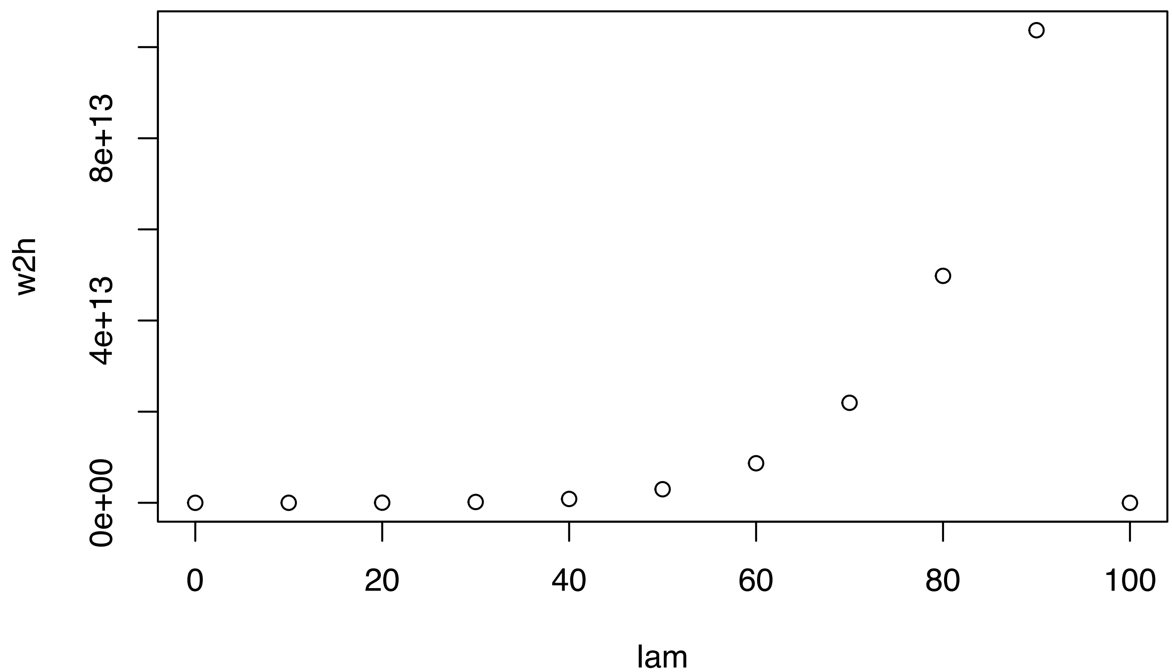
problem 3

c

```
x <- rnorm(100, 0,1)
e <- rnorm(100,0,0.1)

ti <- 1 + 0.2*x -1*x^2 +e


wl1 <- function(lam1,l){
    w1l <- e%*%e+lam1*sqrt(sum(w1*w1))
    w1l
}

wl2 <- function(lam2,l){
   w2l <-  e%*%e+lam2*(sum(w2*w2))
   w2l
}

w1h <- rep(0,11)
w2h <- rep(0,11)
h=0
lam = c(0,10,20,30,40,50,60,70,80,90,100)
for( i in lam){
    w1 = 0.2
    w2 = 1
    for(k in 1: 3){
      w1 <- wl1(lam1 = i,l=k)
      w2 <- wl2(lam2 = i,l=k)
    }
    w1h[h] = w1
    w2h[h] = w2
    h=h+1
}

plot(lam,w1h)
```




```
plot(lam,w2h)
```



d)

w1,2 is increasing when lambda1,2 increasing

problem 6

d)

```
library("readxl")
train<-read_excel("training.xlsx",col_names = c("x1","x2","group"))
fit1 <- glm(group~ x1+x2,data=train)
summary(fit1)
```

```
##
## Call:
## glm(formula = group ~ x1 + x2, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.76184  -0.15292   0.00143   0.17241   0.63974
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.194053   0.043108  50.897   <2e-16 ***
## x1          -0.214846   0.007725 -27.813   <2e-16 ***
## x2           0.017735   0.014143   1.254    0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.06226525)
##
##     Null deviance: 66.667  on 299  degrees of freedom
## Residual deviance: 18.493  on 297  degrees of freedom
## AIC: 23.442
##
## Number of Fisher Scoring iterations: 2
```

```
fit2 <- glm(group~ x1+x2-1,data = train)
summary(fit2)
```

```
##
## Call:
## glm(formula = group ~ x1 + x2 - 1, data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6618  -0.2639   0.1928   0.7601   1.9616
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## x1 -0.03808    0.02148  -1.773   0.0772 .
## x2  0.60288    0.02564  23.514   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.6033199)
##
##     Null deviance: 900.00  on 300  degrees of freedom
## Residual deviance: 179.79  on 298  degrees of freedom
## AIC: 703.76
##
## Number of Fisher Scoring iterations: 2
```

e)

from the coding in d), AIC for model 1 is 23.442, AIC for model 2 is 703.76

f)

```
trdata1 <- predict(fit1, newdata = train, type = "response")
trdata2 <- predict(fit2, newdata = train, type = "response")
test<-read_excel("test.xlsx",col_names = c("x1","x2","group"))
```

```r
testdata1 <- predict(fit1, newdata = test, type = "response" )
testdata2 <- predict(fit2, newdata = test, type = "response")
## training model 1
glm.pred1=rep(1,300)
glm.pred1[trdata1 >1.5]=2
table(glm.pred1,factor(train$group))
```

```
##
## glm.pred1   1   2
##         1  90   5
##         2  10 195
```

```r
## training model 2
glm.pred2=rep(1,300)
glm.pred2[trdata2 >1.5]=2
table(glm.pred2,factor(train$group))
```

```
##
## glm.pred2   1   2
##         1  62 103
##         2  38  97
```

```r
## test model 1
glm.test1 = rep(1,40)
glm.test1[testdata1 >1.5]=2
table(glm.test1,factor(test$group))
```

```
##
## glm.test1  1  2
##         1 19  1
##         2  1 19
```

```r
## test model 2
glm.test2=rep(1,40)
glm.test2[testdata2 >1.5]=2
table(glm.test2,factor(test$group))
```

```
##
## glm.test2  1  2
##         1 12 12
##         2  8  8
```