STAT-S676

Assignment 1

John Koo

See source code here: https://github.com/johneverettkoo/stats-hw

Problem 1

AIC is given by $-2\log(L(\hat{\theta}|y)) + 2K$ where L is the likelihood function. We are given that $Y \sim \mathcal{N}(X\beta, \sigma^2 I)$. Then $\varepsilon = Y - X\beta \sim (0, \sigma^2 I)$. Therefore, the pdf of ε is a multivariate normal $f(\varepsilon) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\varepsilon^T\varepsilon}$. $\varepsilon^T\varepsilon = RSS = n\hat{\sigma}^2 \approx n\sigma^2$. (We sub in the estimate for the parameter.) Then we get $L = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{n}{2}}$. Plugging this in for L, we get:

$$
\begin{aligned}
AIC &= -2\log\left(\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{n}{2}}\right) + 2K \\
&= -2\bigg(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{n}{2}\bigg) + 2K \\
&= n\log(2\pi) + n\log(\sigma^2) + n + 2K
\end{aligned}
$$

We can ignore the two terms $n\log(2\pi)$ and n since they are fixed for some particular dataset.

Problem 2

We have the following components for y, β, and σ^2:

$$f(y|\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{(y-X\beta)^T(y-X\beta)}{2\sigma^2}}$$

$$f(\beta|\sigma^2) = \frac{\left|X^T X\right|^{1/2}}{(2\pi n\sigma^2)^{p/2}} e^{-\frac{(X\beta)^T(X\beta)}{2n\sigma^2}}$$

$$f_{\sigma^2}(\sigma^2) = \left(\frac{1}{2\pi(\sigma^2)^3}\right)^{1/2} e^{-\frac{1}{2\sigma^2}}$$

Then

$$f(y, \beta, \sigma^2) = (y|\beta, \sigma^2) \times f(\beta|\sigma^2) \times f_{\sigma^2}(\sigma^2)$$

And to compute the marginal for Y:

$$f_Y(y) = \int_{\beta,\sigma^2} f(y, \beta, \sigma^2) d\beta d\sigma^2$$

The strategy is to integrate out y and β by completing the square. Combining the exponents, we obtain:

$$\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\left(\frac{\left|X^TX\right|}{(2\pi n\sigma^2)^p}\right)^{1/2}\left(\frac{1}{2\pi(\sigma^2)^3}\right)^{1/2}\exp\left(-\frac{1}{2\sigma^2}\Big((y-X\beta)^T(y-X\beta)+\frac{1}{n}\beta^TX^TX\beta+1\Big)\right)$$

Then rearranging the terms inside the exponent:

$$-\frac{1}{2\sigma^2}\left((\beta-\hat{\beta}\frac{n}{n+1})^T\big(X^TX(1+\frac{1}{n})\big)\big(\beta-\hat{\beta}\frac{n}{n+1}\big)+y^T\big(I-\frac{n}{n+1}H\big)y+1\right)$$

Where $\hat{\beta}=(X^TX)^{-1}X^Ty$ and $H=X(X^TX)^{-1}X^T$.

Then we can see that $\Sigma_\beta=\sigma^2\frac{n}{n+1}(X^TX)^{-1}$. Then integrating w.r.t. β, we get a factor of $\big((2\pi)^p(\sigma^2)^p(\frac{n}{n+1})^p/\left|X^TX\right|\big)^{1/2}$, which leaves us with:

$$f_{Y,\sigma^2}=\left(\frac{1}{2\pi\sigma^2}\right)^{n/2+1/2}\left(\frac{1}{n+1}\right)^{p/2}(\sigma^2)^{-\frac{3}{2}-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}\Big(y^T\big(I-\frac{n}{n+1}H\big)y+1\Big)\right)$$

Integrating out σ^2, we obtain:

$$\left(\frac{1}{2\pi}\right)^{\frac{n+1}{2}}\left(\frac{1}{n+1}\right)^{p/2}\frac{\Gamma(\frac{n+1}{2})}{\left(\frac{1}{2}y^T(I-\frac{n}{n+1}H)y+1/2\right)^{\frac{n+1}{2}}}$$

Problem 3

TIC is defined by $-2\log L+2tr(J(\theta)I(\theta^{-1}))$. $L(\theta|y)=g(y|\theta)$ where $I=\nabla_\theta\nabla_\theta^Tl$ and $J=(\nabla_\theta)^T(\nabla_\theta)$. In terms of partial derivatives, this becomes:

$$I=-E_{truth}\begin{bmatrix}\frac{\partial^2l}{\partial\beta\partial\beta^T} & \frac{\partial^2l}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2l}{\partial\sigma^2\partial\beta^T} & \frac{\partial^2l}{\partial(\sigma^2)^2}\end{bmatrix}$$

$$J=E_{truth}\begin{bmatrix}(\partial_\beta l)(\partial_\beta l)^T & (\partial_\beta l)(\partial_{\sigma^2}l)^T\\ (\partial_{\sigma^2}l)(\partial_{\beta^2}l)^T & (\partial_{\sigma^2}l)(\partial_{\sigma^2}l)^T\end{bmatrix}$$

Then computing the partial derivatives:

$$\frac{\partial l}{\partial\beta}=\frac{(y-X\beta)^TX}{\sigma^2}$$

$$\frac{\partial l}{\partial\sigma^2}=-\frac{1}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}(y-X\beta)^T(y-X\beta)$$

$$\frac{\partial^2l}{\partial\beta\beta^T}=-\frac{X^TX}{\sigma^2}$$

$$\frac{\partial^2l}{\partial\beta\sigma^2}=-\frac{(y-X\beta)^TX}{(\sigma^2)^2}$$

$$\frac{\partial^2l}{\partial(\sigma^2)^2}=\frac{1}{2(\sigma^2)^2}-\frac{1}{(\sigma^2)^3}(y-X\beta)^T(y-X\beta)$$

Noting that $y\sim\mathcal{N}(\mu,\sigma^2I)$, taking the expectations, we arrive at:

$$I = \frac{1}{\sigma^2}\begin{bmatrix} X^TX & \frac{X^T(\mu - X\beta)}{\sigma^2} \\ \frac{(\mu - X\beta)^TX}{\sigma^2} & -\frac{1}{2\sigma^2} + \frac{\sigma^2 + (\mu - X\beta)^T(\mu - X\beta)}{(\sigma^2)^2} \end{bmatrix}$$

Where the numerator of the second term of $[I]_{22}$ was derived as follows: $(y - X\beta)^T(y - X\beta) = (y - \mu + \mu - X\beta)^T(y - \mu + \mu - X\beta) = (y - \mu)^T(y - \mu) + (\mu - X\beta)^T(\mu - X\beta) + 2(y - \mu)^T(\mu - X\beta)$. Then taking the expectation of this, the first term becomes σ^2, the second term stays the same, and the last term goes to 0 since $E[y] = \mu$.

In order to compute J, we need the following:

$$[J]_{11} = E\Big[\frac{1}{(\sigma^2)^2}X^T(y - X\beta)(y - X\beta)^TX\Big]$$

The employing the same trick as before, the middle part becomes:

$$\begin{gathered}(y - \mu + \mu - X\beta)(y - \mu + \mu - X\beta)^T \\ = (y - \mu)(y - \mu)^T + (y - \mu)(\mu - X\beta)^T + (\mu - X\beta)(y - \mu)^T + (\mu - X\beta)(\mu - X\beta)^T\end{gathered}$$

Under the expectation, the middle two terms go to 0 since $E[y] = \mu$. The first term also turns into $\sigma^2 I$. Then we get:

$$[J]_{11} = \frac{1}{(\sigma^2)^2}X^T(\sigma^2 I + (\mu - X\beta)(\mu - X\beta)^T)X$$

For $[J]_{22}$, we compute $\left(\frac{\partial^2 l}{\partial(\sigma^2)^2}\right)^2$ then take the expected value. We can say that the odd powers go to zero under the expectation since the normal is symmetric. Then we get:

$$\frac{1}{4(\sigma^2)^2}\left(1 - \frac{2}{\sigma^2}\big(\sigma^2 + (\mu - X\beta)^2\big) + \frac{1}{(\sigma^2)^2}\big(3(\sigma^2)^2 + 6\sigma^2(\mu - X\beta)^T(\mu - X\beta) + ((\mu - X\beta)^T(\mu - X\beta)^T)^2\big)\right)$$

For the $[J]_{12} = (\frac{\partial l}{\partial \beta})(\frac{\partial l}{\partial \sigma^2})^T$ term:

$$(\frac{\partial l}{\partial \beta})(\frac{\partial l}{\partial \sigma^2})^T = \frac{1}{2(\sigma^2)^2}(y - X\beta)^TX\bigg(-1 + \frac{1}{\sigma^2}(y - X\beta)^T(y - X\beta)\bigg)$$

Taking the expectation of this *should* yield:

$$\frac{1}{2(\sigma^2)^2}\bigg((\mu - X\beta) + \frac{1}{\sigma^2}(3\sigma^2(\mu - X\beta) + (\mu - X\beta)^T(\mu - X\beta)(\mu - X\beta)^T\bigg)$$

$[J]_{21}$ is just the transpose of $[J]_{12}$.

Problem 4

From the textbook, we get the TIC for $tr(JI^{-1})$ assuming $Y - X\beta \sim \mathcal{N}(0, \sigma^2 I)$:

$$tr(JI^{-1}) = \frac{\sigma^2}{\hat{\sigma}^2}\left(K + 1 + \frac{\sigma^2}{\hat{\sigma}^2}(\frac{\gamma - 1}{2} - 2)\right)$$

Where $\gamma = \frac{E[\varepsilon^4]}{E[\varepsilon^2]^2}$. Since we don't know the parameter σ^2, we will just sub in the estimate. Then this turns into:

$$K + \frac{\gamma - 1}{2} - 1$$

Note that when we assume that the errors are normal, $\gamma = 3$, and we just get back the AIC (thanks, textbook).

This is nowhere near "efficient" ...

```
# packages, etc.
import::from(magrittr, `%>%`, `%<>%`)
dp <- loadNamespace('dplyr')
import::from(purrr, flatten)
import::from(foreach, foreach, `%dopar%`)
library(ggplot2)
theme_set(theme_bw())
import::from(parallel, mclapply, detectCores)

# mc stuff
options(mc.cores = detectCores())

# tic function
# no documentation b/c lazy
# but it takes a lm output and returns a number
TIC <- function(linear.model) {
  # find the log likelihood estimate
  l <- as.numeric(logLik(linear.model))

  # number of beta parameters
  K <- length(linear.model$coefficients)

  # gamma from data
  g <- mean(linear.model$residuals ** 4) /
    mean(linear.model$residuals ** 2) ** 2

  # return TIC
  -2 * l + 2 * (K + (g - 1) / 2) - 1
}

# get the data
load('~/dev/stats-hw/stat-s676/diabetes2.Rdata')

# "sex^2" is a linear combination of other variables
# just going to remove it
diabetes2 %<>% dp$select(-`sex^2`)
```

```r
# vector of predictors
predictors <- diabetes2 %>%
  dp$select(-y) %>%
  colnames()

# list of predictor sets
predictor.sets <- lapply(seq(length(predictors)), function(i) {
  combn(predictors, i, simplify = FALSE)
}) %>%
  flatten()

t0 <- Sys.time()
# data frame of results
out.df <- mclapply(predictor.sets, function(p) {
  # subset to predictor set
  temp.df <- diabetes2 %>%
    dp$select(c('y', p))

  # create the model
  temp.mod <- lm(y ~ ., data = temp.df)

  # compute TIC
  tic <- TIC(temp.mod)

  # compute AIC
  aic <- AIC(temp.mod)

  # compute BIC
  bic <- BIC(temp.mod)

  # combine into data frame
  dplyr::data_frame(AIC = aic,
                    BIC = bic,
                    TIC = tic,
                    p = length(temp.mod$coefficients))
}) %>%
  dp$bind_rows()
Sys.time() - t0
```

```
Time difference of 30.65993 mins
```

```r
# predictor set with lowest TIC
predictor.sets[[which.min(out.df$TIC)]]
```

```
 [1] "sex"   "bmi"   "map"   "tc"    "ldl"   "hdl"   "ltg"   "age^2"
 [9] "bmi^2" "ltg^2" "glu^2"
```

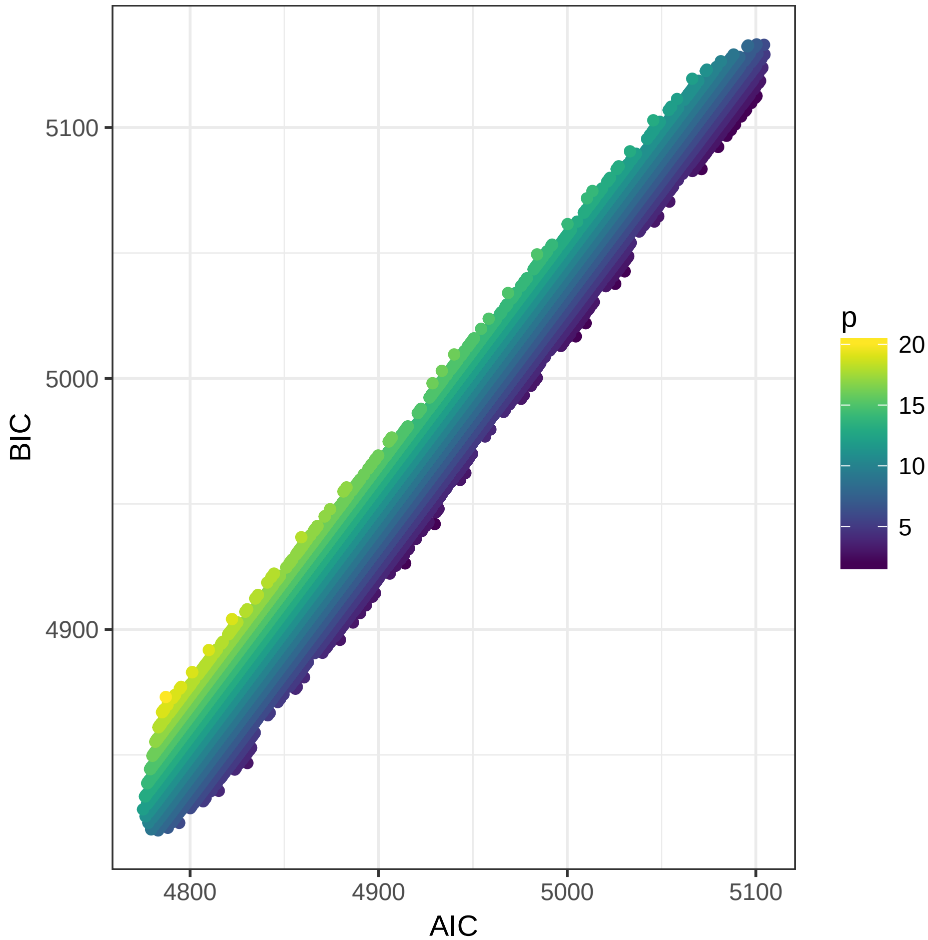
```r
# predictor set with lowest AIC
predictor.sets[[which.min(out.df$AIC)]]
```

```
 [1] "sex"   "bmi"   "map"   "tc"    "ldl"   "hdl"   "ltg"   "age^2"
 [9] "bmi^2" "ltg^2" "glu^2"
```

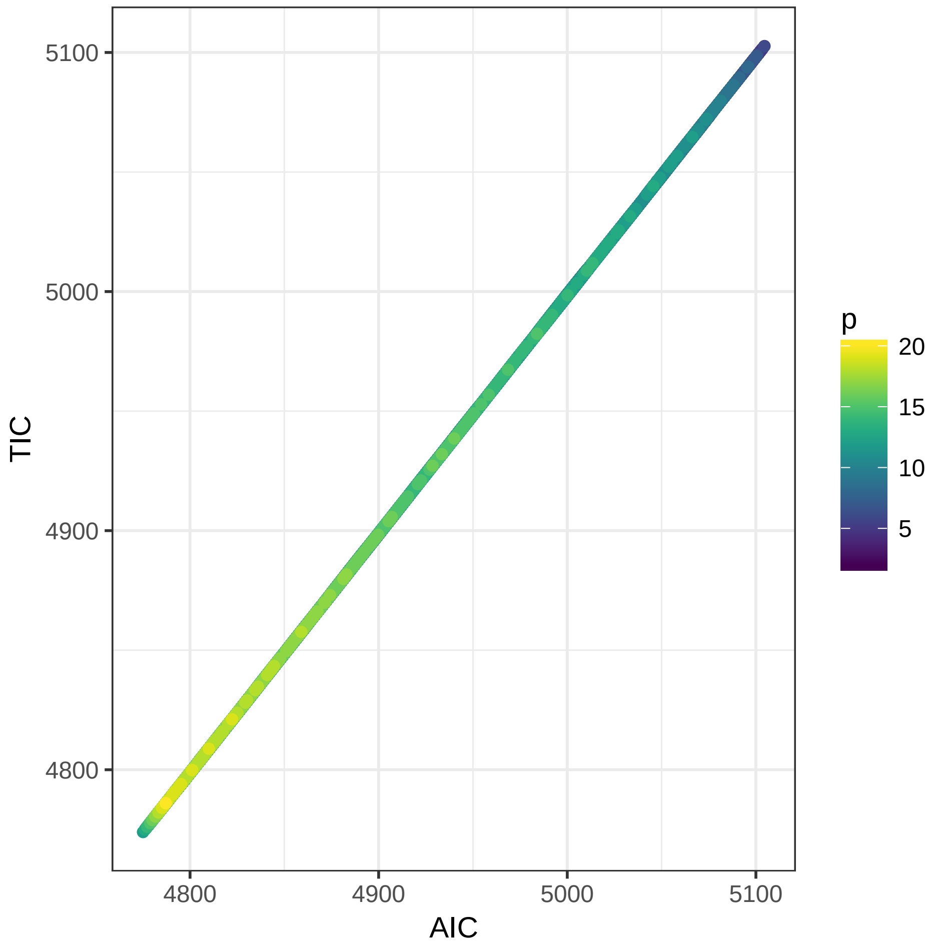
```r
# predictor set with lowest BIC
predictor.sets[[which.min(out.df$BIC)]]
```

```
[1] "sex"    "bmi"    "map"    "tc"     "ldl"    "ltg"    "glu^2"
```

```r
# probably can't see much from this
ggplot(out.df) +
  geom_point(aes(x = AIC, y = BIC, colour = p)) +
  viridis::scale_colour_viridis()
```




```r
ggplot(out.df) +
  geom_point(aes(x = AIC, y = TIC, colour = p)) +
  viridis::scale_colour_viridis()
```


5100
5000
4900
4800
TIC
4800
4900
5000
5100
AIC
p
20
15
10
5