Bayesian Linear Regression: Model Averaging(BMA)

ESC week 9

김정규

Bayesian Model Averaging(BMA)

1. Motivation
   - Model selection scenario
   - Motivating example
   - Advantages of BMA
2. How?
   - Framework
   - Example: MCMC(Gibbs sampler)

Model Selection Scenario

- Several **candidate models** describe data generating process
- However, **uncertainty about model selection process** should be considered

Previous

1. Select the best model (using some criterion)
2. Learn about the parameters of the selected model

BMA

1. Learn parameters of all candidate models
2. Combine estimates according to posterior probability

Motivating Example

- ESC 스터디에 가야 되는데… 버스가 안 온다….
  - 다른 이동수단…지하철? 택시? 킥보드? 무작정 뛰기?

- 확률 모델로 표현하면
  - M_{i}: 이동수단 i를 택했을 때의 평행 우주 (모델 i)
  - $\mathrm{P}(\mathrm{t}|M_{\mathrm{i}})$: 평행우주 i일 때 예상되는 딜레이 t 의 분포
  - $\mathrm{P}(\mathrm{t})$: 예상되는 딜레이의 분포

Motivating Example

- Previous:
  1. 가장 괜찮은 평행 우주 $\hat{M}$ 를 먼저 택한다
  2. $\mathrm{p}(\mathrm{t}|\hat{M})$ 를 통해 결론 (얼마나 늦을지)을 내린다

- BMA:
  1. 모든 평행우주를 동시에 고려한다
  2. $\mathrm{p}(\mathrm{t}) = \sum_{\mathrm{i}} \mathrm{p}(\mathrm{t}|M_{\mathrm{i}})\mathrm{p}(M_{\mathrm{i}})$ 를 통해 얼마나 늦을지 결론을 내린다

Previous vs. BMA

Previous	BMA
평행우주 M_i 의 불확실성을 고려하지 않음 ($p(\hat{M}) = 1$ 이라고 가정)	평행우주 M_i 의 불확실성을 고려 $p(M_{버스}) = 0.6, p(M_{지하철}) = 0.3,$ $p(M_{킥보드}) = 0.08, p(M_{달리기}) = 0.02$
정보 업데이트 불가 (이전에 기각한 새로운 모델을 선택해야 되므로)	정보 업데이트 가능 $P(t\|Data) = \sum_i p(t\|M_i, data)p(M_i\|data)$

- 데이터가 추가됨에 따라 평행우주의 불확실성도 업데이트

Advantages

1. Reduces overconfidence by considering model uncertainty
2. Optimal prediction under several loss function
3. Does not 'reject' model but rather use uncertainty of model for decision making (unlike NHST mentality)
4. Updates posterior estimation as model weights are adjusted
5. Robust to model misspecification

Framework

- 모형의 불확실성을 고려하여 의사결정!

$$\Pr(\Delta \mid Data) = \sum_k \Pr(\Delta|M_k, Data)\Pr(M_k|Data)$$

- Posterior probability of model: $\Pr(M_k|Data) = \frac{\Pr(Data|M_k)\Pr(M_k)}{\sum_l \Pr(Data|M_l)\Pr(M_l)}$
- Marginal likelihood of model: $\Pr(Data|M_k) = \int \Pr(Data|\theta_k, M_k)\Pr(\theta_k|M_k)\,d\theta_k$
- Δ : $quantity\ of\ interest$ *(ex. Future observation, utility of an action…)*
- θ_k: $parameters$

Gibbs Sampler (BMA)

여러 모델($z^{(s)}$)을 샘플링하자!!

$$
\begin{array}{ccccc}
\boldsymbol{z}^{(s)} & \longrightarrow & \sigma^{2(s)} & \longrightarrow & \boldsymbol{\beta}^{(s)} \\
\downarrow & & & & \\
\boldsymbol{z}^{(s+1)} & \longrightarrow & \sigma^{2(s+1)} & \longrightarrow & \boldsymbol{\beta}^{(s+1)}
\end{array}
$$

Algorithm (Gibbs Sampler)

1. Set $\boldsymbol{z} = \boldsymbol{z}^{(s)}$;
2. For $j \in \{1, \ldots, p\}$ in random order, replace z_j with a sample from $p(z_j|\boldsymbol{z}_{-j}, \boldsymbol{y}, \mathbf{X})$;
3. Set $\boldsymbol{z}^{(s+1)} = \boldsymbol{z}$;
4. Sample $\sigma^{2(s+1)} \sim p(\sigma^2|\boldsymbol{z}^{(s+1)}, \boldsymbol{y}, \mathbf{X})$;
5. Sample $\boldsymbol{\beta}^{(s+1)} \sim p(\boldsymbol{\beta}|\boldsymbol{z}^{(s+1)}, \sigma^{2(s+1)}, \boldsymbol{y}, \mathbf{X})$.

Gibbs Sampler (BMA)

Example: Diabetes dataset

- X
  - $x_1, .. x_{10}$: 10 baseline variables (main effect)
  - $\binom{10}{2} = 45$ intercation terms $(x_1x_2, ..., x_1x_{10}, ...)$
  - x_j^2 : quadratic terms (omitting $x_2 =$ sex which is binary)
- Y
  - Disease progression

- Summary
  - $n = 442$ (diabetes subject)
  - $p = 10 + 45 + 9 = 64$ (predictor)
  - Standardize → Train/Test split (342:100) → Do regression
  - Predictive error using: $\hat{y}_{test} = X_{test}\hat{\beta}$ vs. y_{test}

- $\hat{\boldsymbol{\beta}}_{BMA}$?

$$\begin{tikzcd}
\boxed{z^{(s)}} \arrow[r] \arrow[d] & \sigma^{2(s)} \arrow[r] & \beta^{(s)} \\
z^{(s+1)} \arrow[r] & \sigma^{2(s+1)} \arrow[r] & \beta^{(s+1)}
\end{tikzcd}$$

Gibbs Sampler (BMA)

Step1. How to sample $\boldsymbol{z}^{(S)} = \left(z_1^{(S)}, \dots, z_p^{(S)}\right)$?

- $z_j \sim p\left(z_j \,\middle|\, \boldsymbol{y}, \boldsymbol{X}, z_{-j}\right)$
  - z_{-j}: 샘플링 된 j 를 제외한 나머지 regressor
- $z_j = \begin{cases} 1 & w.p. \quad \frac{o_j}{1+o_j} \\ 0 & w.p. \quad \frac{1}{1+o_j} \end{cases} \qquad i.e. \qquad z_j \sim Ber\left(\frac{o_j}{1+o_j}\right)$

$$o_j = \underbrace{\frac{\Pr(z_j = 1|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}{\Pr(z_j = 0|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}}_{\text{Posterior Conditional odds}} = \underbrace{\frac{\Pr(z_j = 1)}{\Pr(z_j = 0)}}_{\text{Prior odds}} \times \underbrace{\frac{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 1)}{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 0)}}_{\text{Bayes factor}}$$

$$
\begin{tikzcd}
z^{(s)} \arrow[r] \arrow[d] & \boxed{\sigma^{2(s)}} \arrow[r] & \beta^{(s)} \\
z^{(s+1)} \arrow[r] & \sigma^{2(s+1)} \arrow[r] & \beta^{(s+1)}
\end{tikzcd}
$$

Gibbs Sampler (BMA)

Step2. How to sample $\sigma^{2(s)}$ & $\boldsymbol{\beta}^{(s)}$?

- $\frac{1}{\sigma^{2(s)}} \sim \text{gamma}\left(\frac{\boldsymbol{\nu}_0 + \text{n}}{2}, \frac{\boldsymbol{\nu}_0 \boldsymbol{\sigma}_0^2 + \text{SSR}_\text{g}^{\text{z}(s)}}{2}\right)$

  - *Note:* $\text{SSR}_g^z = \boldsymbol{y}^T(\mathbf{I} - \frac{g}{g+1}\mathbf{X}_z(\mathbf{X}_z^T\mathbf{X}_z)^{-1}\mathbf{X}_z)\boldsymbol{y}$

Recall

$$
\gamma = 1/\sigma^2 \sim \text{gamma}(\nu_0/2, \nu_0\sigma_0^2/2)
$$

$$
\begin{aligned}
p(\gamma|\boldsymbol{y}, \mathbf{X}) &\propto p(\gamma)p(\boldsymbol{y}|\mathbf{X}, \gamma) \\
&\propto \left[\gamma^{\nu_0/2-1}\exp(-\gamma \times \nu_0\sigma_0^2/2)\right] \times \left[\gamma^{n/2}\exp(-\gamma \times \text{SSR}_g/2)\right] \\
&= \gamma^{(\nu_0+n)/2-1}\exp[-\gamma \times (\nu_0\sigma_0^2 + \text{SSR}_g)/2] \\
&\propto \text{dgamma}(\gamma, [\nu_0 + n]/2, [\nu_0\sigma_0^2 + \text{SSR}_g]/2) ,
\end{aligned}
$$

$$\begin{array}{ccccc} z^{(s)} & \longrightarrow & \sigma^{2(s)} & \longrightarrow & \boxed{\beta^{(s)}} \\ \downarrow & & & & \\ z^{(s+1)} & \longrightarrow & \sigma^{2(s+1)} & \longrightarrow & \beta^{(s+1)} \end{array}$$

Gibbs Sampler (BMA)

Step2. How to sample $\sigma^{2(s)}$ & $\boldsymbol{\beta}^{(s)}$?

- $$\boldsymbol{\beta}^{(s)} \sim \mathrm{MVN}\left(\underbrace{\frac{\mathrm{g}}{1+\mathrm{g}}\,\hat{\beta}_{\mathrm{ols}},}_{E[\boldsymbol{\beta}^{(s)}|\boldsymbol{y},\boldsymbol{X}_{\boldsymbol{z}},\boldsymbol{\sigma}^{2(s)}]}\ \underbrace{\frac{\mathrm{g}}{1+\mathrm{g}}\sigma^{2(s)}\left(\mathrm{X}^{\mathrm{T}}_{\mathrm{z}^{(s)}}\,\mathrm{X}_{\mathrm{z}^{(s)}}\right)}_{\boldsymbol{Var}[\boldsymbol{\beta}^{(s)}|\boldsymbol{y},\boldsymbol{X}_{\boldsymbol{z}},\boldsymbol{\sigma}^{2(s)}]}^{-1}\right)$$

- Recall

$$\begin{aligned}
&p(\boldsymbol{\beta}|\boldsymbol{y},\mathbf{X},\sigma^2) \\
&\propto p(\boldsymbol{y}|\mathbf{X},\boldsymbol{\beta},\sigma^2)\times p(\boldsymbol{\beta}) \\
&\propto \exp\{-\tfrac{1}{2}(-2\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{y}/\sigma^2+\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}/\sigma^2)-\tfrac{1}{2}(-2\boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\beta}_0+\boldsymbol{\beta}^T\Sigma_0^{-1}\boldsymbol{\beta})\} \\
&=\exp\{\boldsymbol{\beta}^T(\Sigma_0^{-1}\boldsymbol{\beta}_0+\mathbf{X}^T\boldsymbol{y}/\sigma^2)-\tfrac{1}{2}\boldsymbol{\beta}^T(\Sigma_0^{-1}+\mathbf{X}^T\mathbf{X}/\sigma^2)\boldsymbol{\beta}\}.
\end{aligned}$$

$$\Sigma_0 = k(\mathbf{X}^T\mathbf{X})^{-1} \quad k = g\sigma^2$$

$\rightarrow$

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\beta}|\boldsymbol{y},\mathbf{X},\sigma^2] &= [\mathbf{X}^T\mathbf{X}/(g\sigma^2)+\mathbf{X}^T\mathbf{X}/\sigma^2]^{-1} \\
&= \frac{g}{g+1}\sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \\
\mathrm{E}[\boldsymbol{\beta}|\boldsymbol{y},\mathbf{X},\sigma^2] &= [\mathbf{X}^T\mathbf{X}/(g\sigma^2)+\mathbf{X}^T\mathbf{X}/\sigma^2]^{-1}\mathbf{X}^T\boldsymbol{y}/\sigma^2 \\
&= \frac{g}{g+1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}.
\end{aligned}$$

Gibbs Sampler (BMA)

요약

$$o_j = \frac{\Pr(z_j = 1|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}{\Pr(z_j = 0|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})} = \frac{\Pr(z_j = 1)}{\Pr(z_j = 0)} \times \frac{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 1)}{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 0)}$$

Posterior Conditional odds — **Prior odds** — **Bayes factor**

Step 1. $z_j \sim p\left(z_j \mid \boldsymbol{y}, \boldsymbol{X}, z_{-j}\right)$

- $z_j = \begin{cases} 1 & w.p. \ \frac{o_j}{1+o_j} \\ 0 & w.p. \ \frac{1}{1+o_j} \end{cases} \qquad i.e. \qquad z_j \sim Ber(\frac{o_j}{1+o_j})$

Step 2. $\frac{1}{\sigma^{2(s)}} \sim \mathrm{gamma}(\frac{\boldsymbol{\nu}_0 + \mathrm{n}}{2}, \ \frac{\boldsymbol{\nu}_0 \boldsymbol{\sigma}_0^2 + \mathrm{SSR}_{\mathrm{g}}^{\mathrm{z}^{(s)}}}{2})$

$\boldsymbol{\beta}^{(s)} \sim \mathrm{MVN}\left(\frac{\mathrm{g}}{1+\mathrm{g}} \hat{\beta}_{\mathrm{ols}}, \ \frac{\mathrm{g}}{1+\mathrm{g}} \sigma^{2(s)} \left(\mathrm{X}_{\mathrm{z}^{(s)}}^{\mathrm{T}} \mathrm{X}_{\mathrm{z}^{(s)}}\right)^{-1}\right)$

우리의 목표:

$$\widehat{\boldsymbol{\beta}}_{BMA} = \frac{\sum_{s=1}^{S} \boldsymbol{\beta}^{(s)}}{S}$$

$\hat{\beta}_{BMA}$?

코드를 통해 이해해보자..!

Algorithm (Gibbs Sampler)

Step 1. $z_j \sim p(z_j \mid \boldsymbol{y}, \boldsymbol{X}, z_{-j})$

$$o_j = \frac{\Pr(z_j = 1|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}{\Pr(z_j = 0|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})} = \frac{\Pr(z_j = 1)}{\Pr(z_j = 0)} \times \frac{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 1)}{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 0)}$$

Posterior Conditional odds — Prior odds — Bayes factor

- $z_j = \begin{cases} 1 & w.p. \ \frac{o_j}{1+o_j} \\ 0 & w.p. \ \frac{1}{1+o_j} \end{cases}$ $\quad i.e. \quad z_j \sim Ber(\frac{o_j}{1+o_j})$

Step 2. $\frac{1}{\sigma^{2(s)}} \sim \text{gamma}(\frac{\boldsymbol{\nu}_0 + \text{n}}{2}, \frac{\boldsymbol{\nu}_0 \boldsymbol{\sigma}_0^2 + \text{SSR}_{\text{g}}^{\text{z}^{(s)}}}{2})$

$$\boldsymbol{\beta}^{(s)} \sim \text{MVN}\left(\frac{\text{g}}{1+\text{g}} \hat{\beta}_{\text{ols}}, \frac{\text{g}}{1+\text{g}} \sigma^{2(s)} (\text{X}_{\text{z}^{(s)}}^{\text{T}} \text{X}_{\text{z}^{(s)}})^{-1}\right)$$

MCMC loop

```
p<-dim(X)[2]
S<-10000
BETA<-Z<-matrix(NA,S,p)
z<-rep(1,dim(X)[2] )
lpy.c<-lpy.X(y,X[,z==1,drop=FALSE])
for(s in 1:S)
{
  for(j in sample(1:p))
  {
    zp<-z ; zp[j]<-1-zp[j]
    lpy.p<-lpy.X(y,X[,zp==1,drop=FALSE])
    r<- (lpy.p - lpy.c)*(-1)^(zp[j]==0)
    z[j]<-rbinom(1,1,1/(1+exp(-r)))
    if(z[j]==zp[j]) {lpy.c<-lpy.p}
  }

  beta<-z
  if(sum(z)>0){beta[z==1]<-lm.gprior(y,X[,z==1,drop=FALSE],S=1)$beta }
  Z[s,]<-z
  BETA[s,]<-beta
```

Step1: Calculating $p(z_j|y, X, z_{-j})$

```
lpy.X<-function(y,X,
                g=length(y),nu0=1,s20=try(summary(lm(y~-1+X))$sigma^2,silent=TRUE))
{
  n<-dim(X)[1] ; p<-dim(X)[2]
  if(p==0) { s20<-mean(y^2) }
  H0<-0 ; if(p>0) { H0<- (g/(g+1)) * X%*%solve(t(X)%*%X)%*%t(X) }
  SS0<- t(y)%*%( diag(1,nrow=n)  - H0 ) %*%y

  -.5*n*log(2*pi) +lgamma(.5*(nu0+n)) - lgamma(.5*nu0)  - .5*p*log(1+g) +
   .5*nu0*log(.5*nu0*s20) -.5*(nu0+n)*log(.5*(nu0*s20+SS0))
}
```

Step 2: Sampling $\sigma^{2^{(s)}}, \beta^{(s)}$

```
lm.gprior<-function(y,X,g=dim(X)[1],nu0=1,
                    s20=try(summary(lm(y~-1+X))$sigma^2,silent=TRUE),S=1000)
{
  n<-dim(X)[1] ; p<-dim(X)[2]
  Hg<- (g/(g+1)) * X%*%solve(t(X)%*%X)%*%t(X)
  SSRg<- t(y)%*%( diag(1,nrow=n)  - Hg ) %*%y

  s2<-1/rgamma(S, (nu0+n)/2, (nu0*s20+SSRg)/2 )

  Vb<- g*solve(t(X)%*%X)/(g+1)
  Eb<- Vb%*%t(X)%*%y

  E<-matrix(rnorm(S*p,0,sqrt(s2)),S,p)
  beta<-t(  t(E%*%chol(Vb)) +c(Eb))

  list(beta=beta,s2=s2)
}
```

Step1: Calculating $p(z_j|y, X, z_{-j})$

```r
p<-dim(X)[2]
S<-10000
BETA<-Z<-matrix(NA,S,p)
z<-rep(1,dim(X)[2] )
lpy.c<-lpy.X(y,X[,z==1,drop=FALSE])
for(s in 1:S)
{
  for(j in sample(1:p))
  {
    zp<-z ; zp[j]<-1-zp[j]
    lpy.p<-lpy.X(y,X[,zp==1,drop=FALSE])
    r<- (lpy.p - lpy.c)*(-1)^(zp[j]==0)
    z[j]<-rbinom(1,1,1/(1+exp(-r)))
    if(z[j]==zp[j]) {lpy.c<-lpy.p}
  }

  beta<-z
  if(sum(z)>0){beta[z==1]<-lm.gprior(y,X[,z==1,drop=FALSE],S=1)$beta }
  Z[s,]<-z
  BETA[s,]<-beta
```

```r
lpy.X<-function(y,X,
               g=length(y),nu0=1,s20=try(summary(lm(y~-1+X))$sigma^2,silent=TRUE))
{
  n<-dim(X)[1] ; p<-dim(X)[2]
  if(p==0) { s20<-mean(y^2) }
  H0<-0 ; if(p>0) { H0<- (g/(g+1)) * X%*%solve(t(X)%*%X)%*%t(X) }
  SS0<- t(y)%*%( diag(1,nrow=n)  - H0 ) %*%y

  -.5*n*log(2*pi) +lgamma(.5*(nu0+n)) - lgamma(.5*nu0)  - .5*p*log(1+g) +
   .5*nu0*log(.5*nu0*s20) -.5*(nu0+n)*log(.5*(nu0*s20+SS0))
}
```

Algorithm (Gibbs Sampler)

Step 1. $z_j \sim p\left(z_j \,\middle|\, \boldsymbol{y}, \boldsymbol{X}, z_{-j}\right)$

- $z_j = \begin{cases} 1 & w.p. \ \frac{o_j}{1+o_j} \\ 0 & w.p. \ \frac{1}{1+o_j} \end{cases} \quad i.e. \quad z_j \sim Ber\left(\frac{o_j}{1+o_j}\right)$

$$o_j = \underbrace{\frac{\Pr(z_j = 1|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}{\Pr(z_j = 0|\boldsymbol{y}, \mathbf{X}, \boldsymbol{z}_{-j})}}_{\text{Posterior Conditional odds}} = \underbrace{\frac{\Pr(z_j = 1)}{\Pr(z_j = 0)}}_{\text{Prior odds}} \times \underbrace{\frac{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 1)}{p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}_{-j}, z_j = 0)}}_{\text{Bayes factor}}$$

Step 2. $\frac{1}{\sigma^{2(s)}} \sim \text{gamma}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR_g^{z(s)}}{2}\right)$

$$\boldsymbol{\beta}^{(s)} \sim \text{MVN}\left(\frac{g}{1+g}\hat{\beta}_{ols}, \frac{g}{1+g}\sigma^{2(s)}\left(X_{z^{(s)}}^T X_{z^{(s)}}\right)^{-1}\right)$$

Step 2. Sample $\sigma^{2(s)}$ & $\boldsymbol{\beta}^{(s)}$?

```
for(s in 1:S)
{
  for(j in sample(1:p))
  {
    zp<-z ; zp[j]<-1-zp[j]
    lpy.p<-lpy.X(y,X[,zp==1,drop=FALSE])
    r<- (lpy.p - lpy.c)*(-1)^(zp[j]==0)
    z[j]<-rbinom(1,1,1/(1+exp(-r)))
    if(z[j]==zp[j]) {lpy.c<-lpy.p}
  }

  beta<-z
  if(sum(z)>0){beta[z==1]<-lm.gprior(y,X[,z==1,drop=FALSE],S=1)$beta }
  Z[s,]<-z
  BETA[s,]<-beta
```

```
lm.gprior<-function(y,X,g=dim(X)[1],nu0=1,
                    s20=try(summary(lm(y~-1+X))$sigma^2,silent=TRUE),S=1000)
{
  n<-dim(X)[1] ; p<-dim(X)[2]
  Hg<- (g/(g+1)) * X%*%solve(t(X)%*%X)%*%t(X)
  SSRg<- t(y)%*%( diag(1,nrow=n)  - Hg ) %*%y

  s2<-1/rgamma(S, (nu0+n)/2, (nu0*s20+SSRg)/2 )

  Vb<- g*solve(t(X)%*%X)/(g+1)
  Eb<- Vb%*%t(X)%*%y

  E<-matrix(rnorm(S*p,0,sqrt(s2)),S,p)
  beta<-t(  t(E%*%chol(Vb)) +c(Eb))

  list(beta=beta,s2=s2)
}
```

$$\frac{1}{\sigma^{2(s)}} \sim \text{gamma}\left(\frac{\boldsymbol{\nu}_0 + \text{n}}{2}, \frac{\boldsymbol{\nu}_0\boldsymbol{\sigma}_0^2 + \text{SSR}_{\text{g}}^{\text{z}(s)}}{2}\right)$$

$$\boldsymbol{\beta}^{(s)} \sim \text{MVN}\left(\frac{\text{g}}{1+\text{g}}\,\hat{\beta}_{\text{ols}}, \frac{\text{g}}{1+\text{g}}\,\sigma^{2(s)}\left(\text{X}_{\text{z}(s)}^{\text{T}}\,\text{X}_{\text{z}(s)}\right)^{-1}\right)$$

Algorithm (Gibbs Sampler)

Step 1. $z_j \sim p(z_j \mid \boldsymbol{y}, \boldsymbol{X}, z_{-j})$

$$o_j = \frac{\Pr(z_j=1|\boldsymbol{y},\mathbf{X},\boldsymbol{z}_{-j})}{\Pr(z_j=0|\boldsymbol{y},\mathbf{X},\boldsymbol{z}_{-j})} = \frac{\Pr(z_j=1)}{\Pr(z_j=0)} \times \frac{p(\boldsymbol{y}|\mathbf{X},\boldsymbol{z}_{-j},z_j=1)}{p(\boldsymbol{y}|\mathbf{X},\boldsymbol{z}_{-j},z_j=0)}$$

Posterior Conditional odds — Prior odds — Bayes factor

- $z_j = \begin{cases} 1 & w.p.\ \frac{o_j}{1+o_j} \\ 0 & w.p.\ \frac{1}{1+o_j} \end{cases}$ $\quad i.e.\quad z_j \sim Ber\left(\frac{o_j}{1+o_j}\right)$

Step 2. $\frac{1}{\sigma^{2(s)}} \sim \text{gamma}\left(\frac{\boldsymbol{\nu}_0 + \text{n}}{2}, \frac{\boldsymbol{\nu}_0\boldsymbol{\sigma}_0^2 + \text{SSR}_{\text{g}}^{\text{z}(s)}}{2}\right)$

$$\boldsymbol{\beta}^{(s)} \sim \text{MVN}\left(\frac{\text{g}}{1+\text{g}}\,\hat{\beta}_{\text{ols}}, \frac{\text{g}}{1+\text{g}}\,\sigma^{2(s)}\left(\text{X}_{\text{z}(s)}^{\text{T}}\,\text{X}_{\text{z}(s)}\right)^{-1}\right)$$

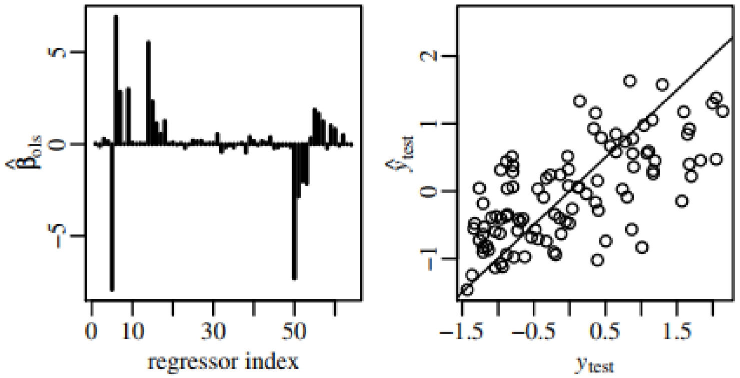
Note on Gibbs sampler for BMA

- Gibbs sampler 에 의해 S 가 충분히 크면
  - $\mathrm{z}^{(\mathrm{s})} \rightarrow \mathrm{p}(\mathrm{z}|\boldsymbol{y}, \boldsymbol{X})$ (모델의 사후분포)
  - $(\sigma^{2(s)}, \beta^{(\mathrm{s})}) \sim \mathrm{p}(\sigma^2, \boldsymbol{\beta}|\boldsymbol{z}^{(s)}, \boldsymbol{y}, \boldsymbol{X}) \rightarrow \mathrm{p}(\sigma^2, \beta|\boldsymbol{y}, \boldsymbol{X})$ (parameter 의 사후분포)
  - $\hat{\beta}_{\mathrm{BMA}} = \frac{\sum_{\mathrm{s}=1}^{\mathrm{S}} \beta^{(\mathrm{s})}}{\mathrm{S}}$

- 가능한 모든 모델: 2^{p}
- 깁스 샘플링에 의해 샘플링된 모델: S 개
  - 한계점: 만약 p 가 너무 크면 posterior approximation 보장 안됨..
  - 그러나 의미 없는 회귀 계수가 많은 경우 reasonable 하다

결과

OLS



BMA

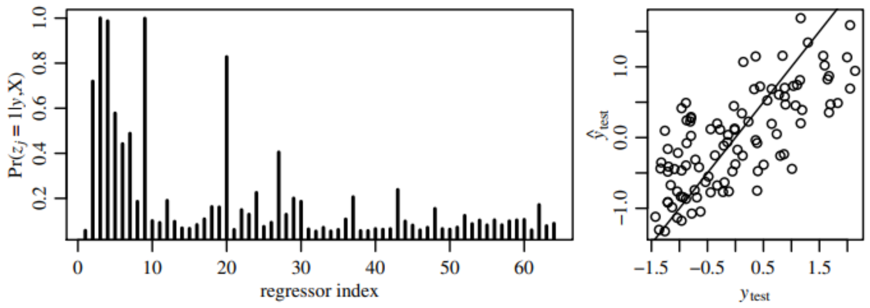


Fig. 9.7. The first panel shows posterior probabilities that each coefficient is non-zero. The second panel shows y_{test} versus predictions based on the model averaged estimate of $\boldsymbol{\beta}$.

- RMSE 가 좋음 (BMA: 0.452 vs. vs. Backward: 0.53, OLS: 0.67)

- $\tilde{y} = P_\pi y$, $\tilde{\boldsymbol{y}} \sim \boldsymbol{X}$? (P_π: random permutation matrix)
    - $\tilde{y}$ 와 X 는 아무런 관계가 없어야 정상… (즉 모든 regressor = 0 을 기대)
    - Backward selection 적용시 18개의 변수가 유의하다고 나온다… !!!
    - BMA 하는 경우 p(**z**)< 0.5 (모델 선택의 불확실성이 반영된 결과!!)