Swati Gupta

Movie Recommendation System

Aim

- Build a movie recommendation system
- Given a set of ratings for movies by users, predict a user’s rating for a movie and provide recommendations based on that
- Use Bayesian techniques and Gibbs sampling
- Data: 96000+ ratings for 6000+ movies by 600+ users

Collaborative Filtering

User preferences are predicted based on the preferences of many people

The basic assumption: similar people will rate similar movies in a similar way

Use latent factors to measure similarity

Properties of a movie that causes a user to like or dislike it like popularity of the cast, duration of the movie, etc.

Model Parameters

L: The number of latent factors

U: The number of users

M: The number of movies

N: The number of observed ratings

Y_{um}: The rating given to movie m by user u

Y: The full vector of observed ratings

γ_m: A vector of length $L+1$ for movie m. The first element $\gamma_m[0]$ is the bias for the movie. The remaining L elements, $\gamma_m[1:]$, are the latent factors associated with the movie

Γ: $M \times (L+1)$ matrix where row m is γ_m

θ_u: A vector of length $L+1$ for user u. The first element $\theta_u[0]$ is the bias for the user. The remaining L elements, $\theta_u[1:]$, are user's preferences for the latent factors

Θ: $U \times (L+1)$ matrix where row u is θ_u

μ: The overall mean of ratings

σ^2: The residual variance of ratings

Model

$$Y_{um} = \mu + \theta_u[0] + \gamma_m[0] + \theta_u[1:]^T \gamma_m[1:] + \epsilon_{um}$$
$$\epsilon_{um} \sim N(0, \sigma^2)$$
$$p(Y_{um}|\mu, \sigma^2, \theta_u, \gamma_m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_{um} - (\mu + \theta_u[0] + \gamma_m[0] + \theta_u[1:]^T \gamma_m[1:]))^2}{2\sigma^2} \right)$$

$$p(\mu) \propto 1$$
$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$
$$p(\gamma_m) = N(0, \Sigma_\gamma)$$
$$p(\theta_u) = N(0, \Sigma_\theta)$$

where Σ_γ, Σ_θ are $L+1 \times L+1$ covariance matrices. They are assumed to be diagonal matrices with all diagonal elements equal to λ_γ and λ_θ respectively both assumed to be 1.

Conditionals

$$p(\mu|Y,\sigma^2,\Theta,\Gamma) = N(\mu_n,\sigma_n^2)$$

$$\text{where } \mu_n = \frac{1}{N}\sum_{i=1}^{N}(y_i - (\theta_{u_i}[0] + \gamma_{m_i}[0] + \theta_{u_i}[1:]^T\gamma_{m_i}[1:]))$$

$$\sigma_n^2 = \frac{\sigma^2}{N}$$

$$p(\sigma^2|Y,\mu,\Theta,\Gamma) = IG(\alpha_n,\beta_n)$$

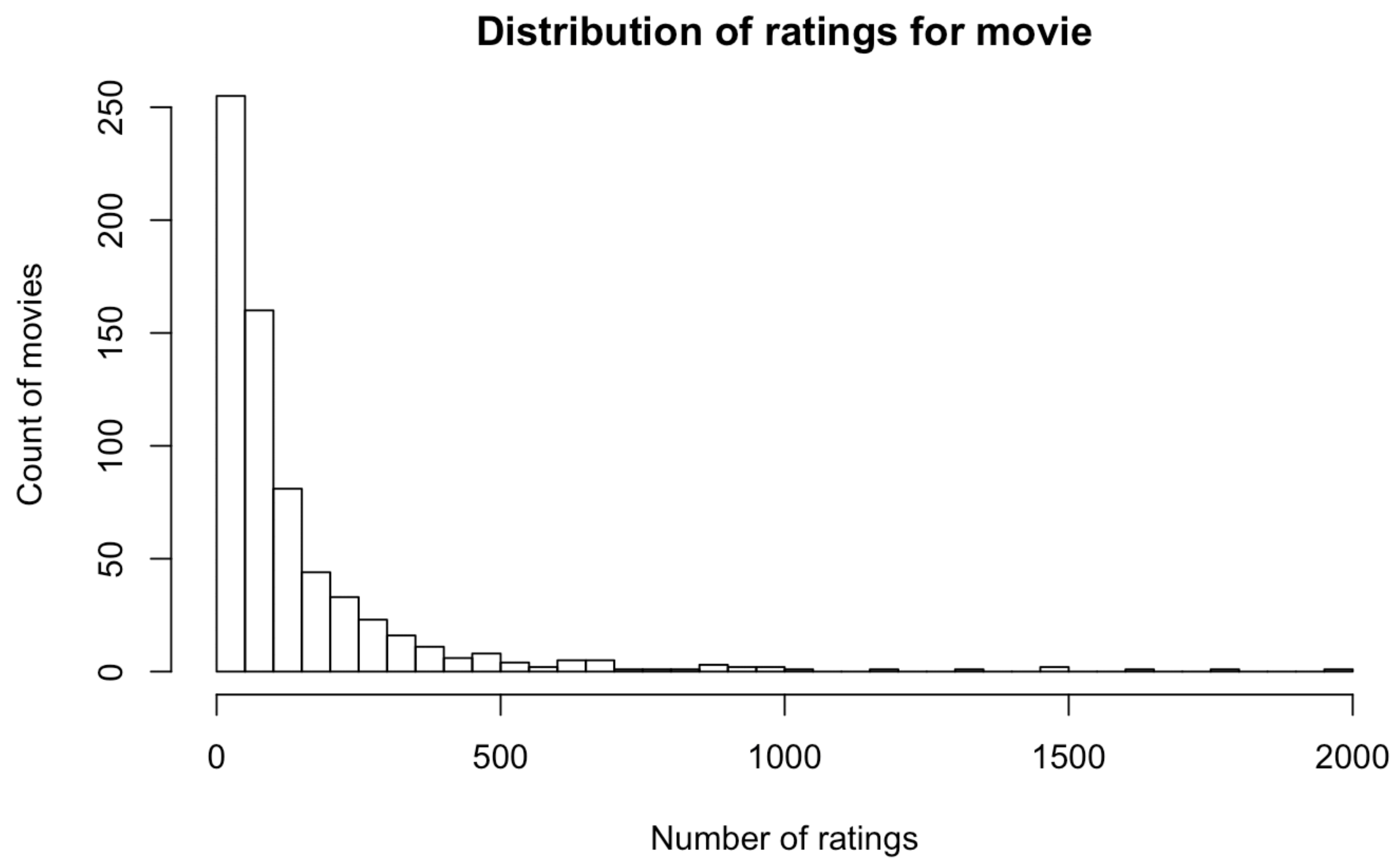
$$\text{where } \alpha_n = \frac{N}{2}$$

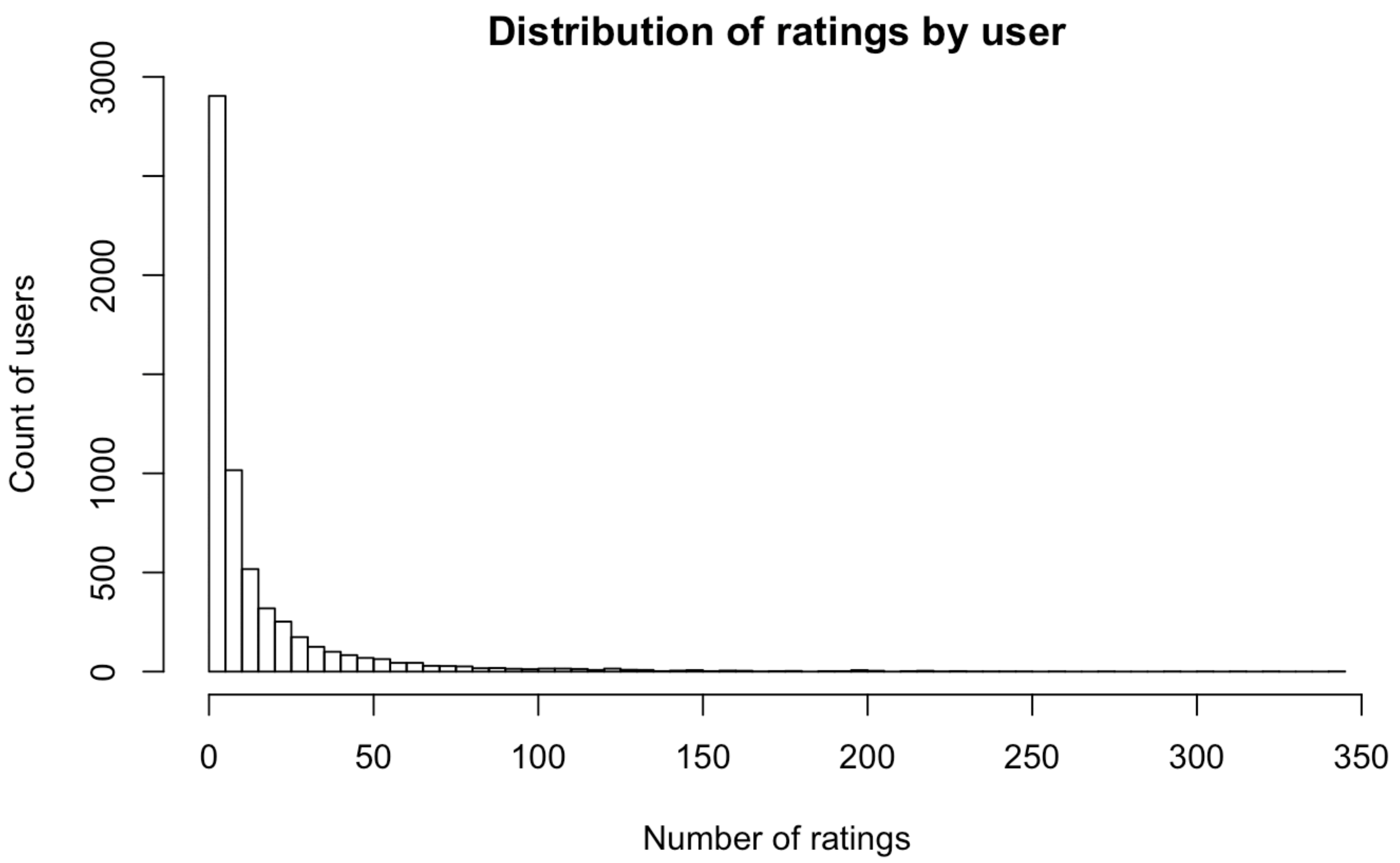
$$\beta_n = \frac{\sum_{i=1}^{N}(y_i - (\mu + \theta_{u_i}[0] + \gamma_{m_i}[0] + \theta_{u_i}[1:]^T\gamma_{m_i}[1:]))^2}{2}$$

Conditionals

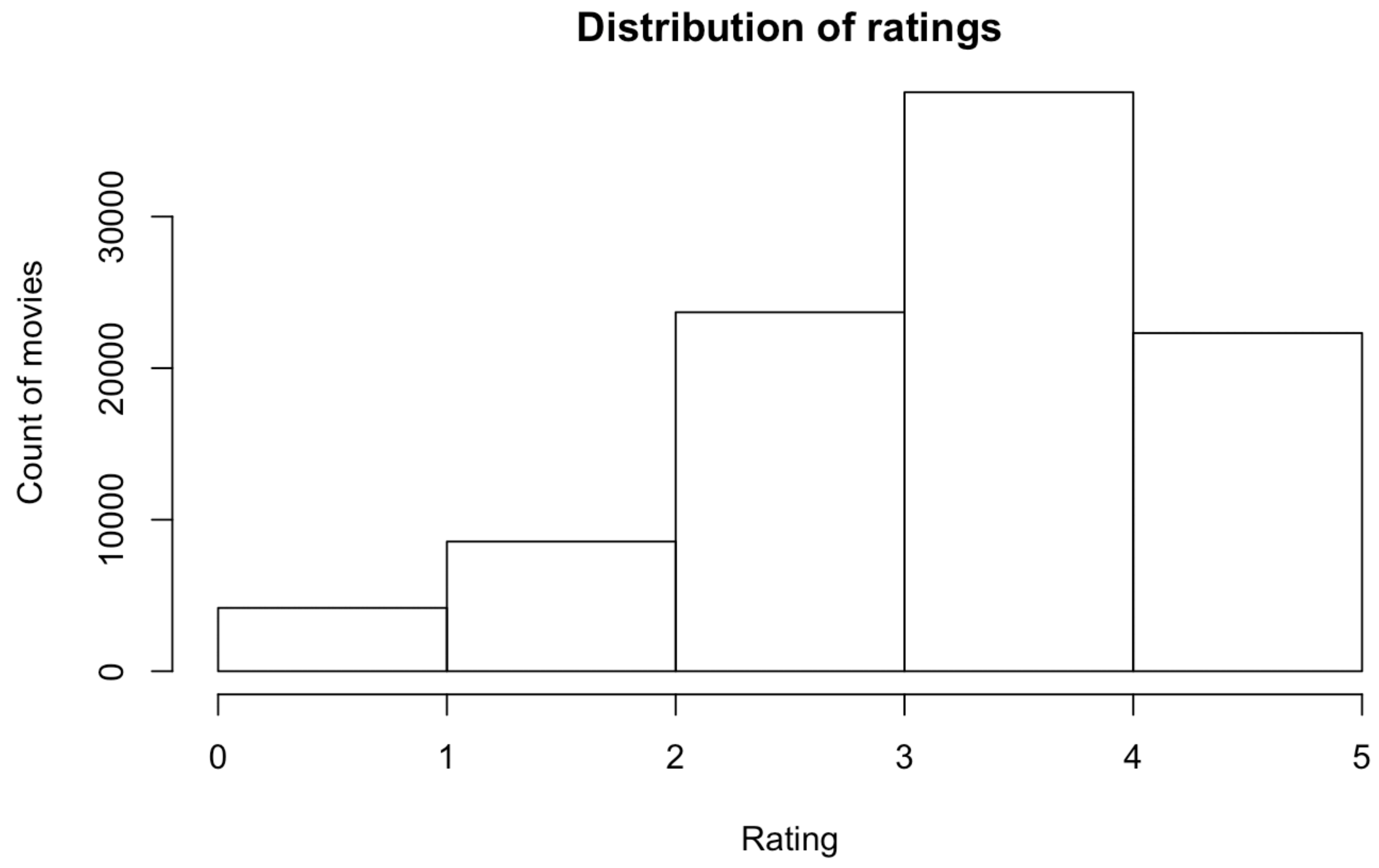
$$
\begin{aligned}
p(\gamma_{m'}|Y, \mu, \sigma^2, \Theta, \Gamma_{-m'}) &= N(\mu_{m'}, \Sigma_{m'}) \\
\text{where } \mu_{m'} &= \Sigma_{m'} \frac{1}{\sigma^2} X_{m'}^T Y_{m'} \\
\Sigma_{m'}^{-1} &= \frac{1}{\sigma^2} X_{m'}^T X_{m'} + \Sigma_{\gamma}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
p(\theta_{u'}|Y, \mu, \sigma^2, \Theta_{-u'}, \Gamma) &= N(\mu_{u'}, \Sigma_{u'}) \\
\text{where } \mu_{u'} &= \Sigma_{u'} \frac{1}{\sigma^2} X_{u'}^T Y_{u'} \\
\Sigma_{u'}^{-1} &= \frac{1}{\sigma^2} X_{u'}^T X_{u'} + \Sigma_{\theta}^{-1}
\end{aligned}
$$


Distribution of ratings for movie
Count of movies
0
50
100
150
200
250
0
500
1000
1500
2000
Number of ratings



Distribution of ratings by user
Count of users
0
500
1000
2000
3000
0
50
100
150
200
250
300
350
Number of ratings


Distribution of ratings
Count of movies
0
10000
20000
30000
0
1
2
3
4
5
Rating

Gibbs Sampling

After removing the burn in samples, the average across all remaining samples was taken to get an estimate of the parameters.

These parameters were then used to predict the ratings in validation set

Initial Values

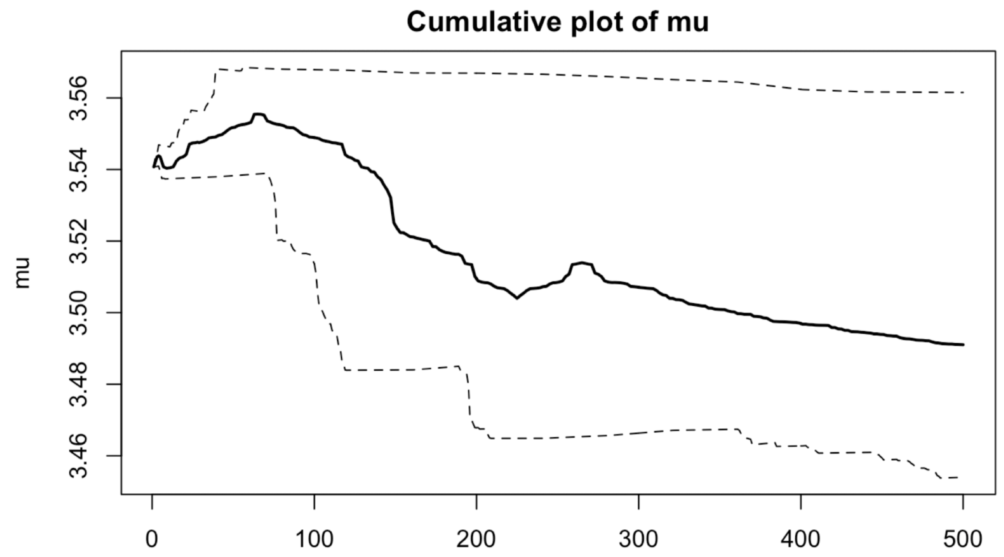
μ: mean of all ratings in training data

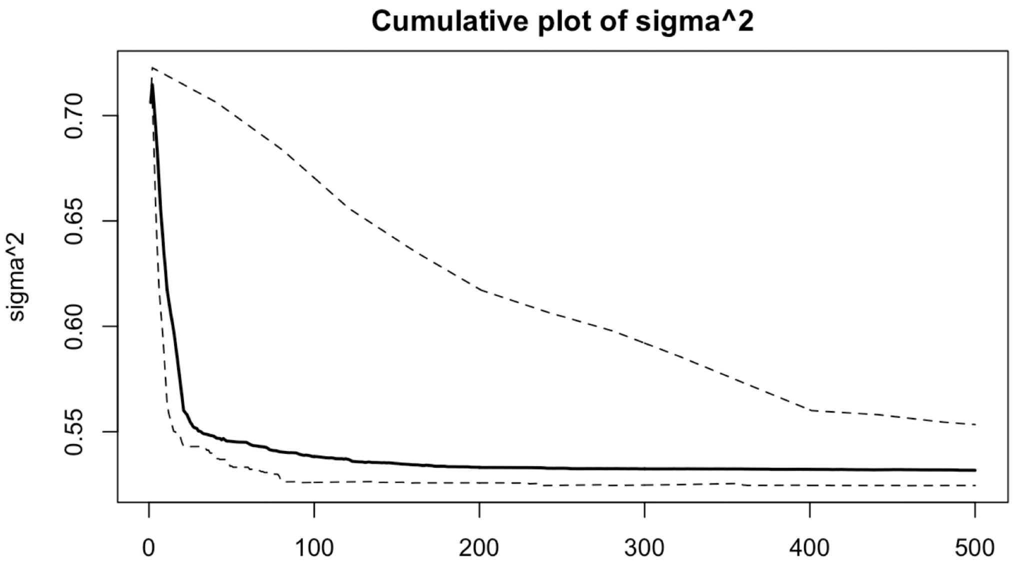
σ^2: variance of all ratings in training data

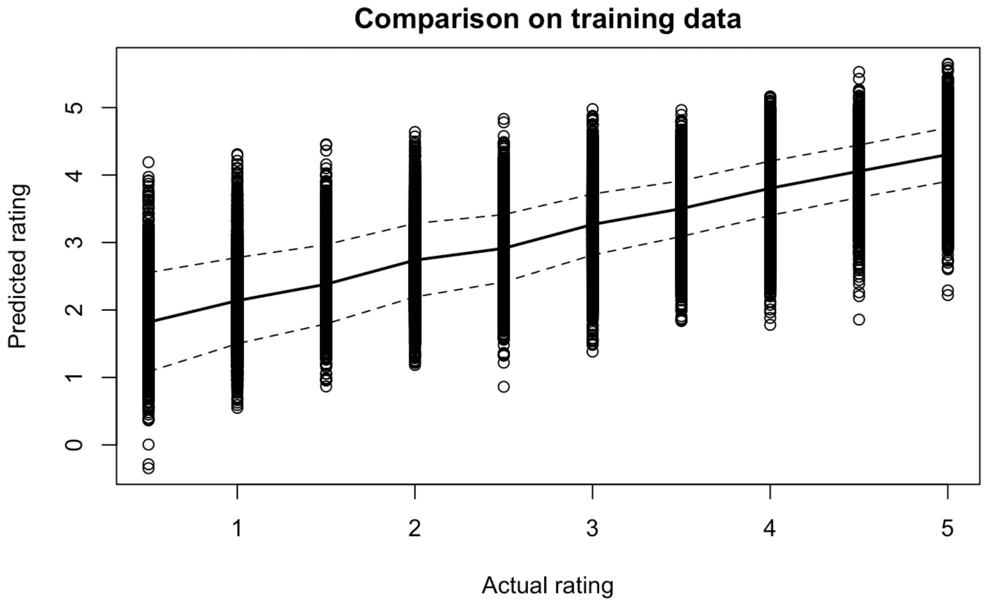
γ_m: $\gamma_m[0] = \mu_m - \mu$ where μ_m is the mean ratings of movie m. $\gamma_m[1:] = 0$

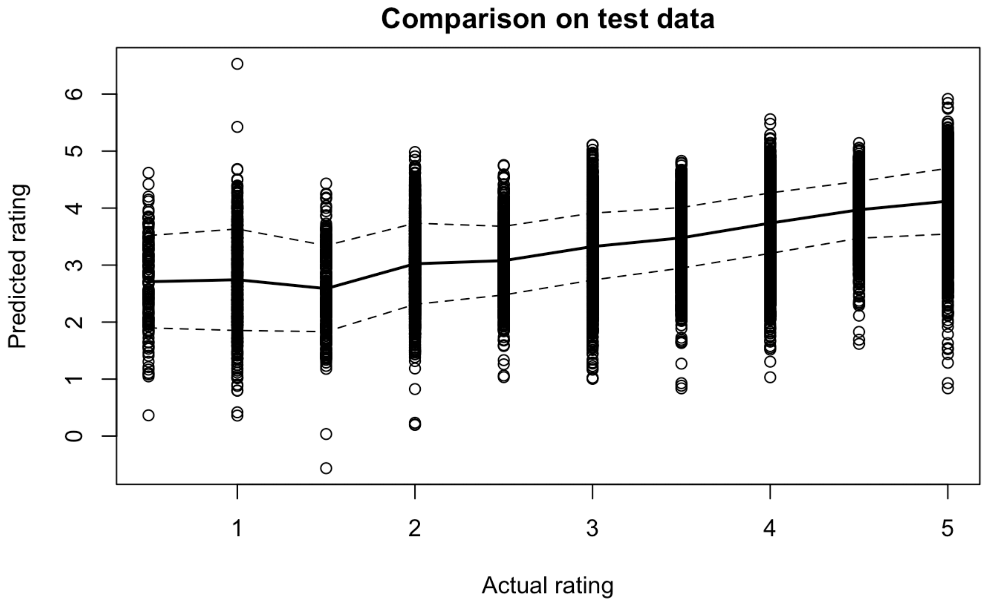
θ_u: $\theta_u[0] = \mu_u - \mu$ where μ_u is the mean ratings by user u. $\theta_u[1:] = 0$

Case 1 : L = 4, number of samples = 500, burn in = 200

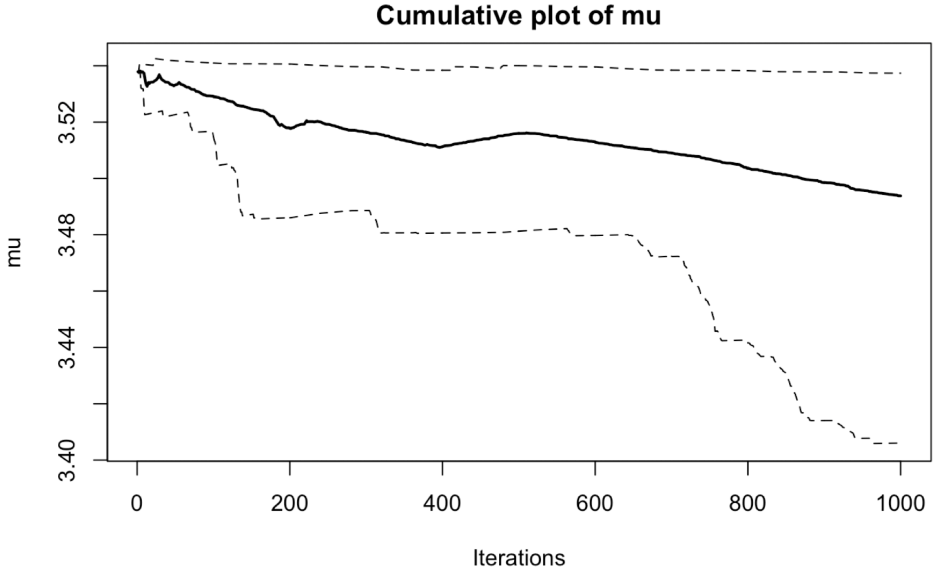


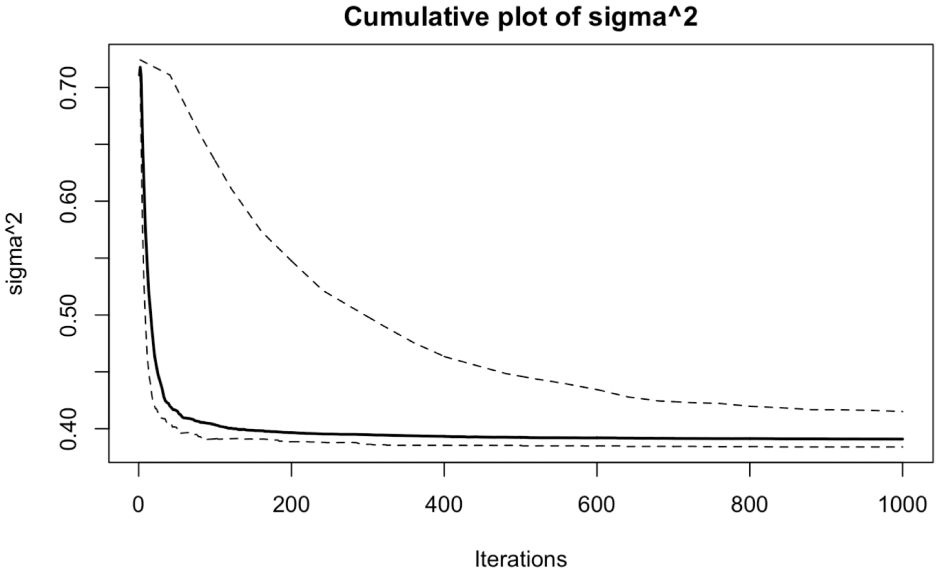


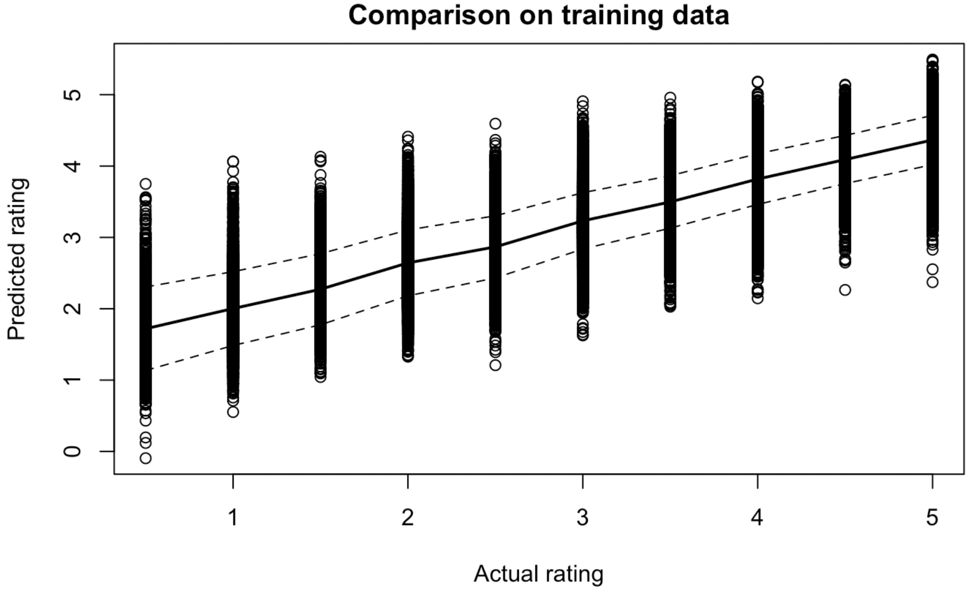


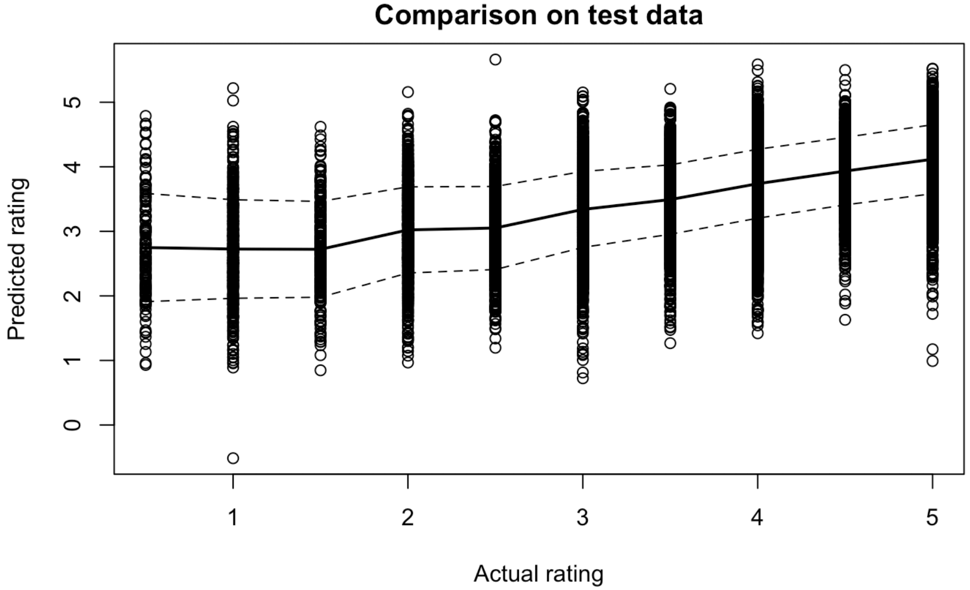


Case 2 : L = 10, number of samples = 1000, burn in = 200

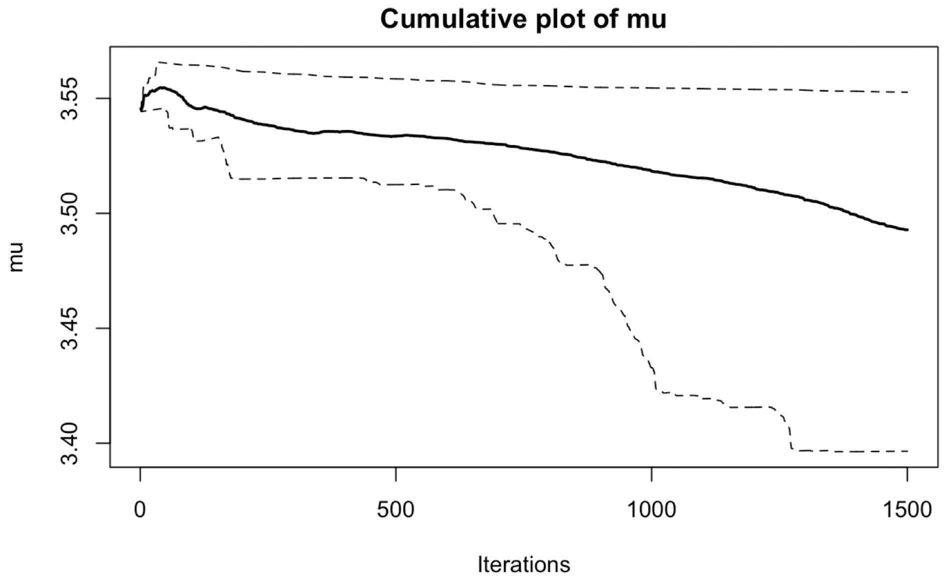


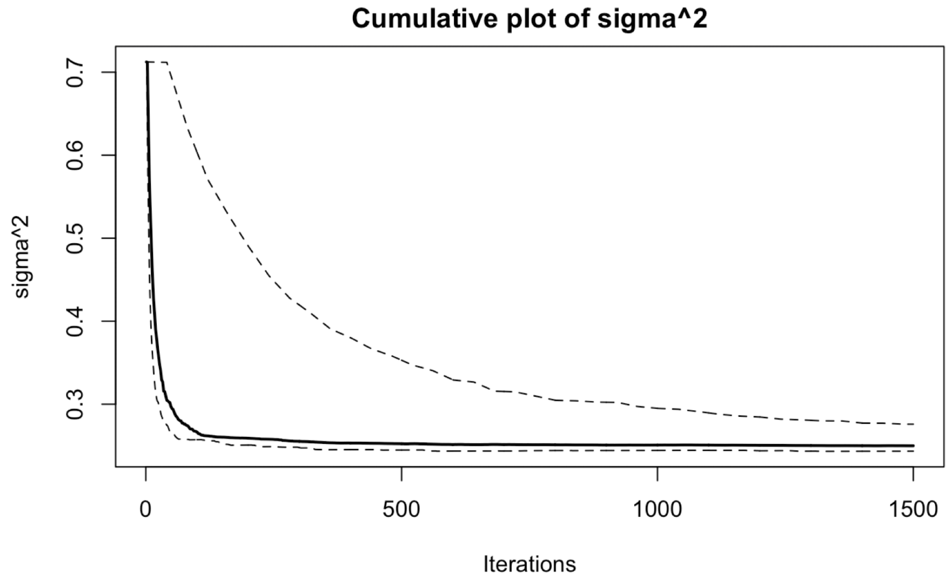


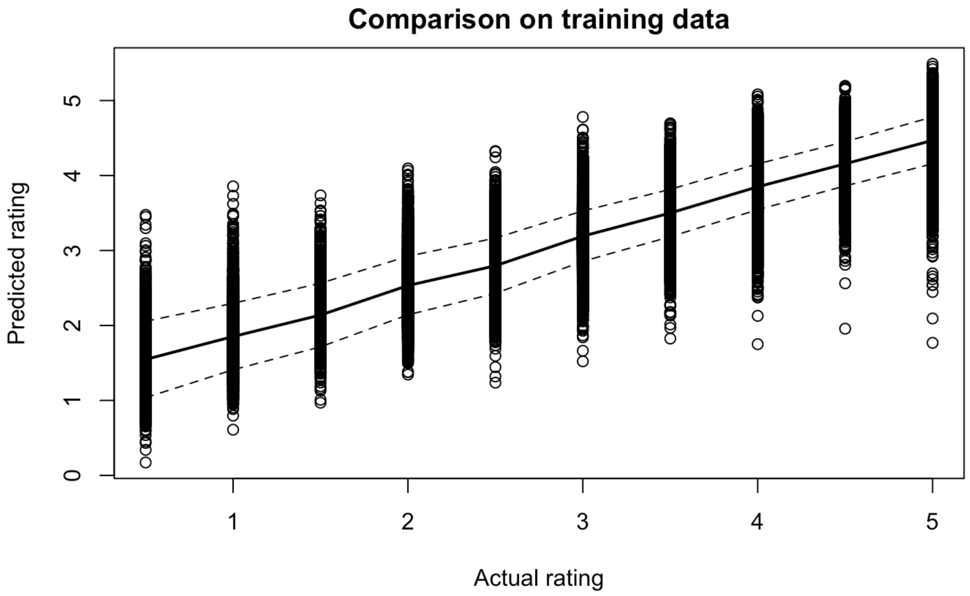


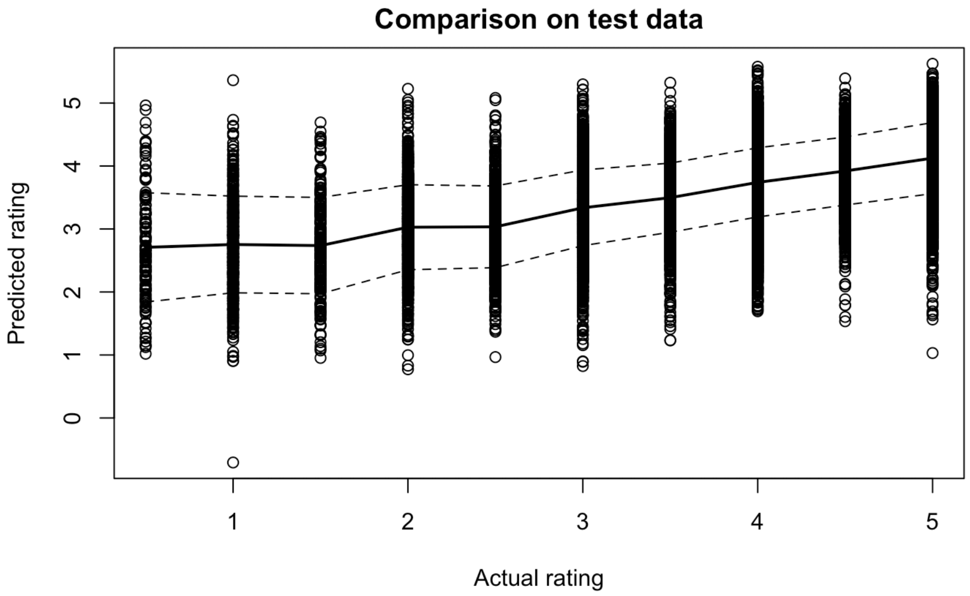


Case 3 : L = 20, number of samples = 1500, burn in = 1000

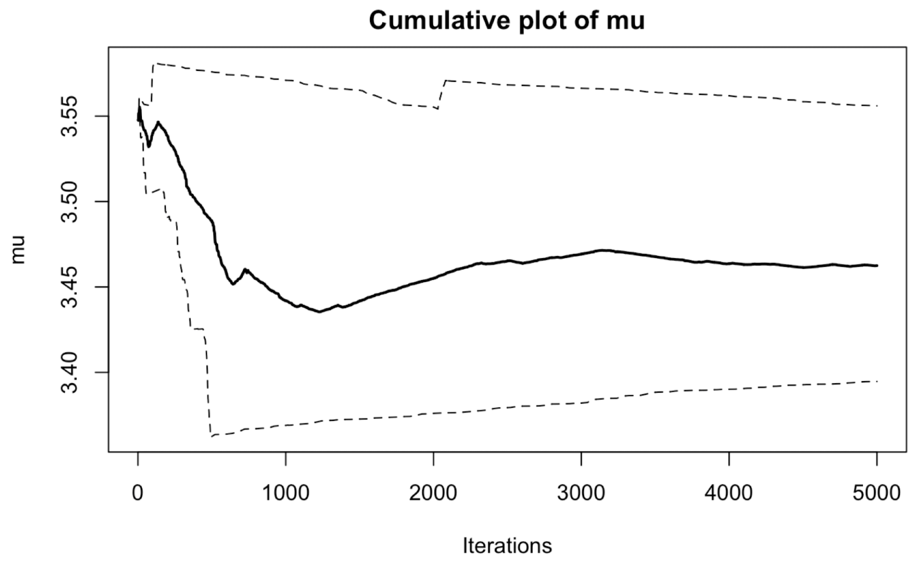


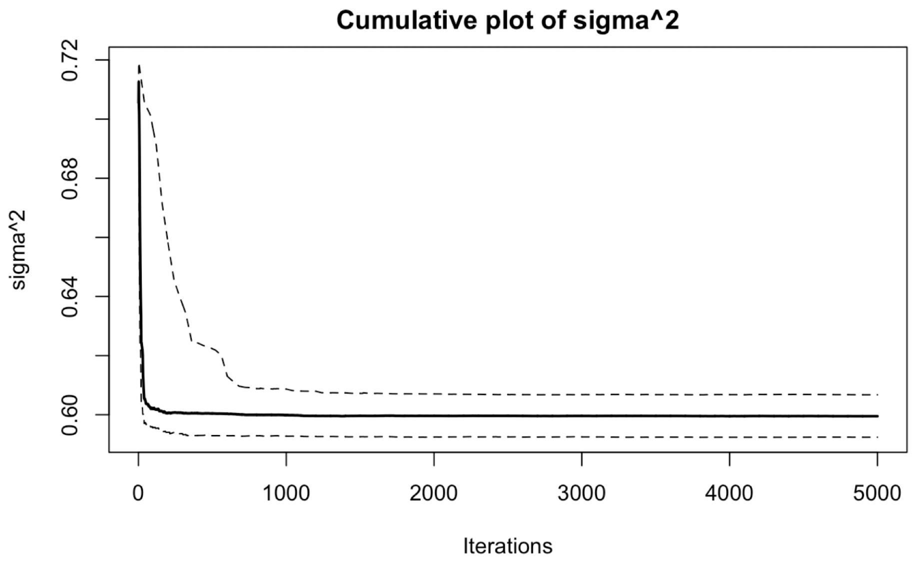


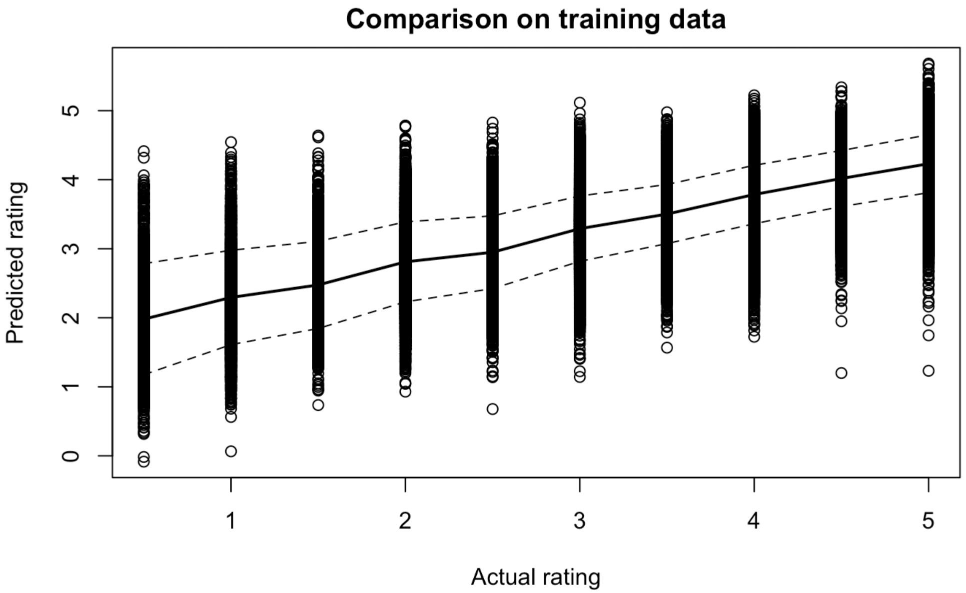


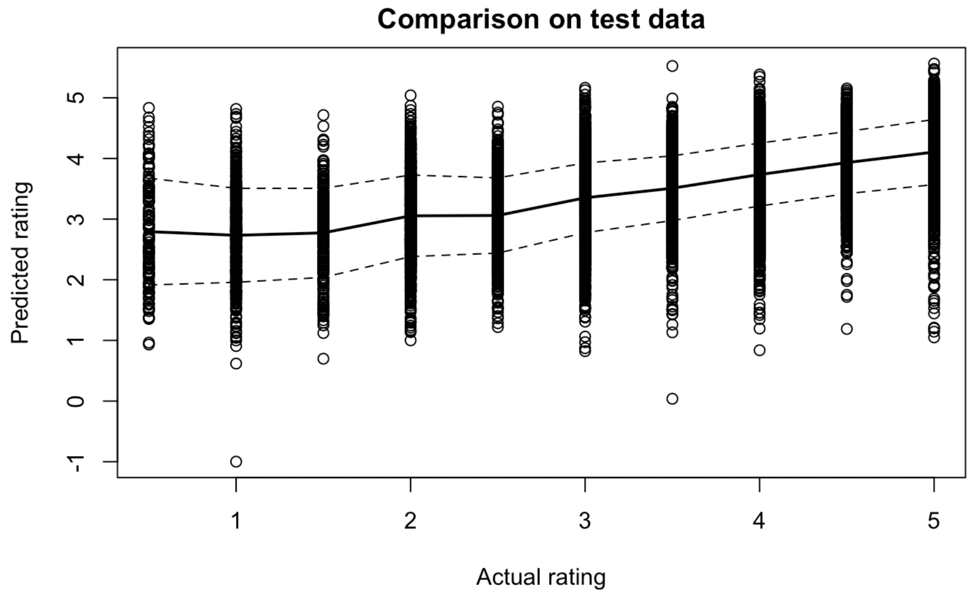


Case 4 : L = 2, number of samples = 5000, burn in = 4000









Comparing models

Case	L	Samples	Burn in	Training error (RMSE)	Validation error (RMSE)
1	4	500	200	0.66	0.87
2	10	1000	200	0.58	0.88
3	20	1500	1000	0.49	0.89
4	2	5000	4000	0.71	0.88

Comparing models

Case	L	Samples	Burn in	Training error (RMSE)	Validation error (RMSE)
1	4	500	200	0.66	0.87
2	10	1000	200	0.58	0.88
3	20	1500	1000	0.49	0.89
4	2	5000	4000	0.71	0.88

Model 4 on test data

μ = 3.46

σ = 0.6

RMSE = 0.86