Exercise 4.1

$p(\beta|X, y) \propto p(y|X, \beta)p(\beta) \sim N(\beta_*, \Sigma_*)$
Given that $f_* = \phi^T \beta$: a linear function. So $f_* \sim N(\phi^T \mathrm{E}(\beta), \phi^T \mathrm{Cov}(\beta)\phi)$

Exercise 4.3

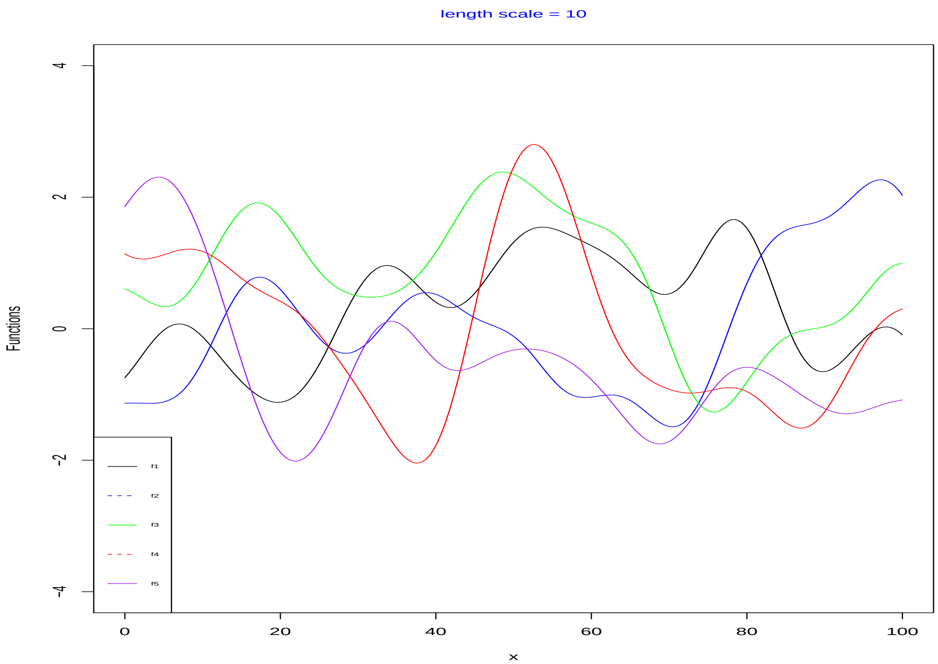

Figure 1: Length scale=10

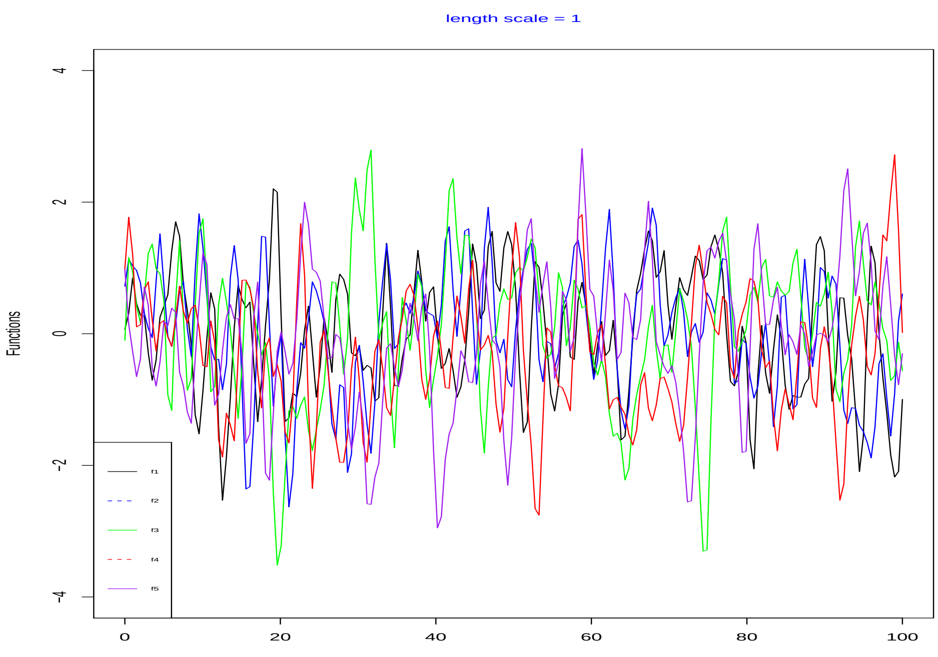

Figure 2: Length scale=1

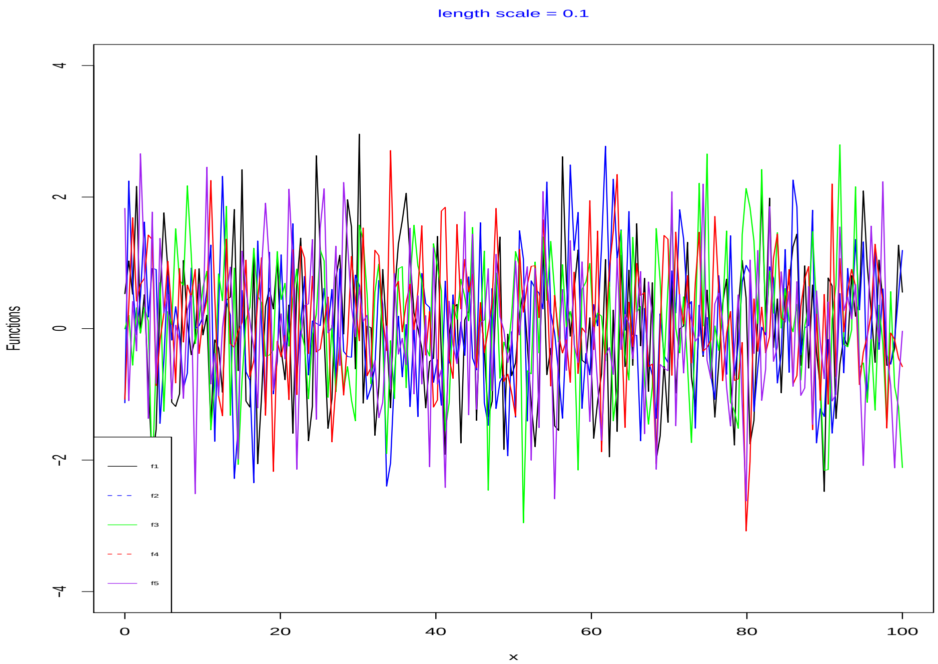

Figure 3: Length scale=0.1

Exercise 4.5

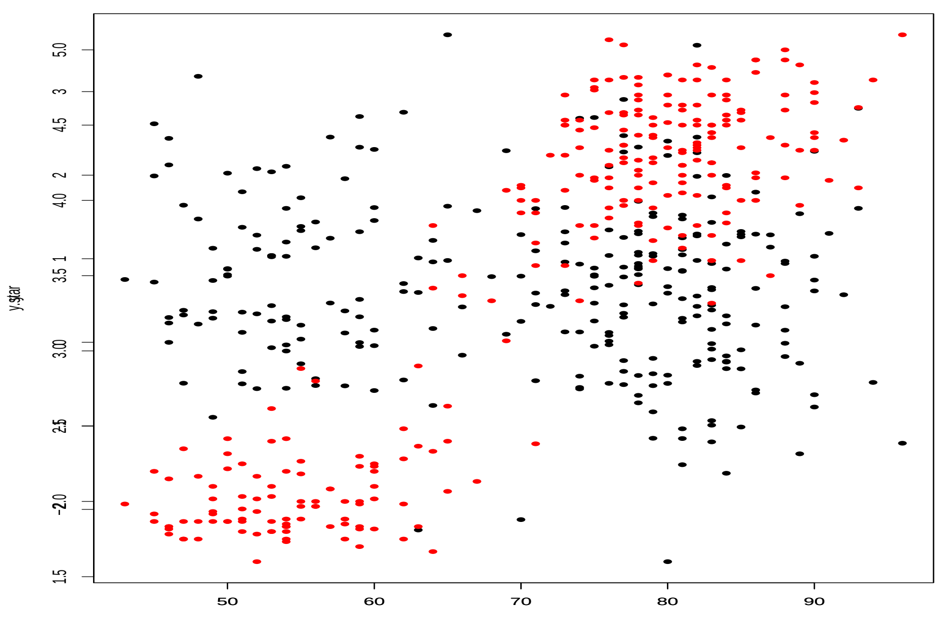

Figure 4: true and predicted values

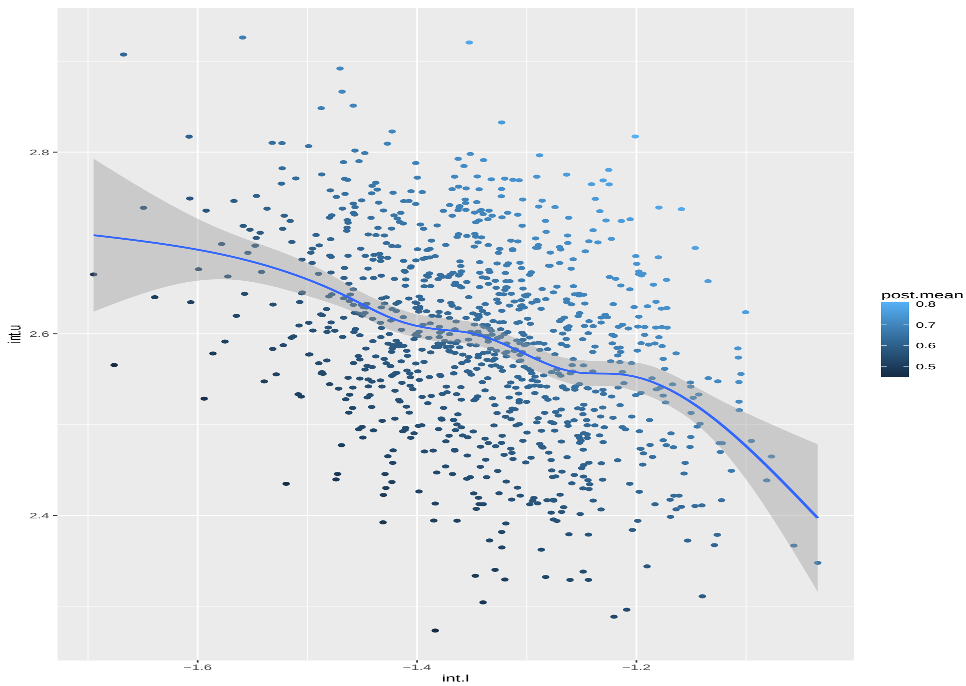

Figure 5: 95% credible intervals

Exercise 4.10

Given a dataset $D = \{(\boldsymbol{x_i}, y_i)|i = 1, .., n\}$, we assume that the labels are generated independently, conditional on $f(x)$. Using the same Gaussian prior $\beta \sim N(0, \Sigma_p)$.
We then obtain the un-normalized posterior: $\log p(\beta|X, y) = -\frac{1}{2}\beta^T\Sigma_p^{-1}\beta + \sum \log(\text{sigmoid}(y_i f_i))$.
For binary classification the basic idea behind Gaussian process prediction is that we place a GP prior over the latent function $f(x)$.
$p(y_i = 1|\boldsymbol{x_i}, \beta) = \frac{1}{1+e^{-x_i^T\beta}} = \frac{1}{1+e^{-f(x_i)}} = \sigma(f(x))$: sigmoid function.
From the reference given in the exercise:
Let $P^*(f) \propto p(f|X, y, \theta)$
$\log P^*(f) = \log p(y|f) + \log p(f|X) = \log p(y|f) - \frac{1}{2}f^T K^{-1} f - \frac{1}{2}\log|K| + \text{const.}$