Homework 2 (4%)
Bayesian Regression

The goal of this homework is to provide you a chance to get familiar with the basic procedure in Bayesian regression. Recall that the posterior distribution of weights and precision is

$$p(w, \beta|\boldsymbol{t})$$

$$= N(w|m_N, \beta^{-1}S_N)Gam(\beta|a_N, b_N)$$

$$m_N = S_N[S_0^{-1}m_0 + \Phi^T\boldsymbol{t}]$$

$$\beta S_N^{-1} = \beta(S_0^{-1} + \Phi^T\Phi)$$

$$a_N = a_0 + \frac{N}{2}$$

$$b_N = b_0 + \frac{1}{2}\left(m_0^T S_0^{-1} m_0 - m_N^T S_N^{-1} m_N + \sum_{n-1}^{N} t_n^2\right)$$

We are interested in seeing how the choice of S_0 influence the prediction accuracy. Split the records in ldpa30_train.csv into two sets. The records from the Week 1 to 270 are Set1 and the remaining records are Set2. Our target variable is week_return13 and our basis function includes week_return1 and week_return4 (and a constant 1). Set $m_0 = 0$, $a_0 = b_0 = 1$. Assume that $S_0 = \lambda I$. Train your model using Set1. Used the learned coefficients to predict Set2. Vary λ and plot the RMSE with respect to λ. Note that to conduct prediction you should:

1. Compute S_N, m_N, a_N, and b_N,
2. Draw β from $Gamma(a_N, b_N)$
3. Draw w from $N(w|m_N, \beta^{-1}S_N)$
4. Use w to make prediction.
5. Compute error.