STAT 153 & 248 - Time Series
Homework Three

Spring 2025, UC Berkeley

Due by 11:59 pm on 10 March 2025
Total Points = 87 (STAT 153) and 105 (STAT 248)

1. Download the Google Trends Data (for the United States) for the query *mask*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025. The goal of this problem is to fit a smooth trend function $\{\mu_t\}$ to the data using the model:

$$y_t = \mu_t + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

and

$$\mu_t = \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \beta_3(t-3)_+ + \cdots + \beta_{n-1}(t-(n-1))_+.$$

   a) Fit the trend function μ_t to the data using ridge regularization with tuning parameter $\lambda = 1, 10, 100$, and 1000. Compare and interpret the differences among these four trend estimates. Which of these estimates provides the best summary for the data? **(4 points)**

   b) Fit the trend function μ_t to the data using LASSO regularization with tuning parameter $\lambda = 1, 10, 100$, and 1000. Compare and interpret the differences among these four trend estimates. Which of these estimates provides the best summary for the data? **(4 points)**

   c) Among the eight estimated trend functions (four from ridge and four from LASSO), which one is the most suitable for this dataset? Explain your reasoning. **(2 points)**

2. Download the FRED dataset on the federal minimum wage hourly wage for nonfarm workers for the United States from `https://fred.stlouisfed.org/series/FEDMINNFRWG`. This is a monthly dataset (from Oct 1938 to Jan 2025). The goal of this problem is to fit a trend function $\{\mu_t\}$ to the data using the model:

$$y_t = \mu_t + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

and

$$\mu_t = \beta_0 + \beta_1 I\{t > 1\} + \beta_2 I\{t > 2\} + \beta_3 I\{t > 3\} + \cdots + \beta_{n-1} I\{t > n-1\}.$$

Here $I\{t > j\}$ equals 1 if $t > j$ and 0 otherwise. Ridge regularization here would use the penalty $\lambda \sum_{j=1}^{n-1} \beta_j^2$ and LASSO regularization would use the penalty $\lambda \sum_{j=1}^{n-1} |\beta_j|$.

a) Fit the trend function μ_t to the data using ridge regularization with tuning parameter $\lambda = 1, 10, 100$, and 1000. Compare and interpret the differences among these four trend estimates. Which of these estimates provides the best summary for the data? **(4 points)**

b) Fit the trend function μ_t to the data using LASSO regularization with tuning parameter $\lambda = 1, 10, 100$, and 1000. Compare and interpret the differences among these four trend estimates. Which of these estimates provides the best summary for the data? **(4 points)**

c) Among the eight estimated trend functions (four from ridge and four from LASSO), which one is the most suitable for this dataset? Explain your reasoning. **(2 points)**

3. Consider a univariate regression dataset $(x_1, y_1), \dots, (x_n, y_n)$ where each x_i and y_i are real-valued. Fix $\lambda > 0$.

a) Suppose $(\hat{\beta}_0^{\text{ridge}}(\lambda), \hat{\beta}_1^{\text{ridge}}(\lambda))$ minimize

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda \beta_1^2$$

over all β_0, β_1. Show that (**3 points**)

$$\hat{\beta}_1^{\text{ridge}}(\lambda) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\lambda + \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0^{\text{ridge}}(\lambda) = \bar{y} - \bar{x}\hat{\beta}_1^{\text{ridge}}(\lambda).$$

b) Suppose $(\hat{\beta}_0^{\text{lasso}}(\lambda), \hat{\beta}_1^{\text{lasso}}(\lambda))$ minimize

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda |\beta_1|$$

over all β_0, β_1. Show that $\hat{\beta}_0^{\text{lasso}}(\lambda) = \bar{y} - \bar{x}\hat{\beta}_1^{\text{lasso}}(\lambda)$ and (**5 Points**)

$$\hat{\beta}_1^{\text{lasso}}(\lambda)) = \begin{cases} \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) - \lambda/2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} & \text{if } \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) > \lambda/2 \\ \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + \lambda/2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} & \text{if } \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) < -\lambda/2 \\ 0 & \text{if } -\lambda/2 \le \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \le \lambda/2. \end{cases}$$

4. Download the Google Trends Data (for the United States) for the query *yahoo*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025. The goal of this problem is to fit a smooth trend function $\{\mu_t\}$ to the data using the model:

$$y_t = \mu_t + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

and

$$\mu_t = \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \beta_3(t-3)_+ + \dots + \beta_{n-1}(t-(n-1))_+.$$

a) Implement ridge regression (with a cross validation strategy for determining the regularization parameter) to estimate μ_t. Report your value of λ and plot the estimated trend function $\hat{\mu}_t$ on a scatter plot of the data. Comment on whether the fitted trend function is capturing well the patterns present in the data without fully interpolating the data. (**6 points**)

b) Consider Bayesian inference with the prior:

$$\beta_0, \beta_1 \overset{\text{i.i.d}}{\sim} N(0, C), \quad \beta_2, \dots, \beta_{n-1} \overset{\text{i.i.d}}{\sim} N(0, \tau^2)$$

along with unif$(-C, C)$ priors for τ and σ (C is a large number).

i. Calculate point estimates $\hat{\tau}$ of τ and $\hat{\sigma}$ of σ. How does your choice of the tuning parameter λ in part (a) compare to $\hat{\sigma}^2/\hat{\tau}^2$? (**4 points**).

ii. Draw $N = 1000$ posterior samples from $\{\mu_t\}$ and plot these sampled trend functions on a scatter plot of the data. Comment on whether the uncertainty revealed in this plot makes sense. (**4 points**)

iii. Average over the N posterior samples to obtain the Bayesian point estimate of $\hat{\mu}_t$. Compare this with the ridge regression estimate from part (a). Which trend estimate is smoother? Which trend estimate would you prefer? (**4 points**).

5. Download the Google Trends Data (for the United States) for the query *golf.* This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025.The goal of this problem is to fit a smooth trend function $\{\mu_t\}$ to the data using the model:

$$y_t = \mu_t + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

and

$$\mu_t = \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \beta_3(t-3)_+ + \cdots + \beta_{n-1}(t-(n-1))_+.$$

a) Implement ridge regression (with a cross validation strategy for determining the regularization parameter) to estimate μ_t. Report your value of λ and plot the estimated trend function $\hat{\mu}_t$ on a scatter plot of the data. Comment on whether the fitted trend function is capturing well the patterns present in the data without fully interpolating the data. (**6 points**)

b) Consider Bayesian inference with the prior:

$$\beta_0, \beta_1 \overset{\text{i.i.d}}{\sim} N(0, C), \quad \beta_2, \dots, \beta_{n-1} \overset{\text{i.i.d}}{\sim} N(0, \tau^2)$$

along with unif$(-C, C)$ priors for τ and σ (C is a large number).

i. Calculate point estimates $\hat{\tau}$ of τ and $\hat{\sigma}$ of σ. How does your choice of the tuning parameter λ in part (a) compare to $\hat{\sigma}^2/\hat{\tau}^2$? (**4 points**).

ii. Draw $N = 1000$ posterior samples from $\{\mu_t\}$ and plot these sampled trend functions on a scatter plot of the data. Comment on whether the uncertainty revealed in this plot makes sense. (**4 points**)

iii. Average over the N posterior samples to obtain the Bayesian point estimate of $\hat{\mu}_t$. Compare this with the ridge regression estimate from part (a). Which trend estimate is smoother? Which trend estimate would you prefer? (**4 points**).

6. Consider again the *golf* google trends dataset as in the previous problem. In this problem, we will consider the model

$$\begin{aligned} y_t &= \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \beta_3(t-3)_+ + \cdots + \beta_{n-1}(t-(n-1))_+ \\ &+ \beta_n \cos\left(2\pi\frac{t}{12}\right) + \beta_{n+1}\sin\left(2\pi\frac{t}{12}\right) + \epsilon_t. \end{aligned}$$

a) A natural ridge regression estimator here would minimize:

$$\sum_{t=1}^{n}\Bigg(y_t - \beta_0 - \beta_1(t-1) - \beta_2(t-2)_+ - \beta_3(t-3)_+ - \cdots - \beta_{n-1}(t-(n-1))_+ \\ - \beta_n \cos\left(2\pi\frac{t}{12}\right) - \beta_{n+1}\sin\left(2\pi\frac{t}{12}\right)\Bigg)^2 + \lambda\sum_{j=2}^{n-1}\beta_j^2$$

Implement this estimator with a cross validation strategy for determining the regularization parameter. Report your value of λ and plot the estimated trend functions:

$$\hat{\mu}_t := \hat{\beta}_0 + \hat{\beta}_1(t-1) + \hat{\beta}_2(t-2)_+ + \hat{\beta}_3(t-3)_+ + \cdots + \hat{\beta}_{n-1}(t-(n-1))_+$$

and

$$\hat{\mu}_t + \hat{\beta}_n \cos\left(2\pi\frac{t}{12}\right) + \hat{\beta}_{n+1}\sin\left(2\pi\frac{t}{12}\right).$$

on a scatter plot of the data. Comment on whether these trend functions capture well the patterns present in the data without fully interpolating the data. (**8 points**)

b) Consider Bayesian inference with the prior:

$$\beta_0, \beta_1, \beta_n, \beta_{n+1} \overset{\text{i.i.d}}{\sim} N(0, C), \quad \beta_2, \ldots, \beta_{n-1} \overset{\text{i.i.d}}{\sim} N(0, \tau^2)$$

along with unif$(-C, C)$ priors for τ and σ (C is a large number).

i. Calculate point estimates $\hat{\tau}$ of τ and $\hat{\sigma}$ of σ. How does your choice of the tuning parameter λ in part (a) compare to $\hat{\sigma}^2/\hat{\tau}^2$? (**5 points**).

ii. Draw $N = 1000$ posterior samples from $\{\mu_t\}$ (note that μ_t does not include the sinusoidal part) and plot these sampled trend functions on a scatter plot of the data. Comment on whether the uncertainty revealed in this plot makes sense. (**5 points**)

iii. Average over the N posterior samples to obtain the Bayesian point estimate of $\hat{\mu}_t$. Compare this with the ridge regression estimate from part (a). Which trend estimate is smoother? Which trend estimate would you prefer? (**5 points**).

7. [**This question is only for students taking STAT 248**] Consider Bayesian inference for the model studied in Lecture 12:

$$y_t = \beta_0 + \beta_1(t-1) + \beta_2(t-2)_+ + \cdots + \beta_{n-1}(t-(n-1))_+ + \epsilon_t \tag{1}$$

with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$, along with the prior:

$$\beta \mid \tau, \sigma \sim N(0, Q), \log\tau \sim \text{uniform}(-C, C), \log\sigma \sim \text{uniform}(-C, C)$$

where Q is the $n \times n$ diagonal matrix with diagonal entries $C, C, \tau^2, \ldots, \tau^2$ (C is as usual a large constant). In Lecture 12, we discussed a gridding-based procedure for Bayesian inference in this model. In this problem, we shall explore the Gibbs sampler for obtaining posterior samples $\beta^{(i)}, \tau^{(i)}, \sigma^{(i)}$ for $i = 1, \ldots, N = 5000$. The Gibbs sampler employs the following algorithm:

a) Initialize $\beta^{(0)}, \tau^{(0)}, \sigma^{(0)}$.

b) For each $i = 1, 2, \ldots, N$,

i. Sample $\beta^{(i)}$ from the conditional posterior distribution of β given $\tau = \tau^{(i-1)}$, $\sigma = \sigma^{(i-1)}$ as well as the data.

ii. Sample $\tau^{(i)}$ from the conditional posterior distribution of τ given $\beta = \beta^{(i)}$, $\sigma = \sigma^{(i-1)}$ as well as the data.

iii. Sample $\sigma^{(i)}$ from the conditional posterior distribution of σ given $\beta = \beta^{(i)}$, $\tau = \tau^{(i)}$ as well as the data.

For the following, you may assume $C \to \infty$ (below X denotes the design matrix corresponding to the regression model (1))

a) Prove that (**3 points**)

$$\beta \mid \text{data}, \sigma, \tau \sim N\left(\left(\frac{X^T X}{\sigma^2} + Q^{-1}\right)^{-1} \frac{X^T y}{\sigma^2}, \left(\frac{X^T X}{\sigma^2} + Q^{-1}\right)^{-1}\right)$$

b) Prove that (**3 points**)

$$\frac{1}{\tau^2} \mid \text{data}, \sigma, \beta \sim \text{Gamma}\left(\frac{n-2}{2}, \frac{1}{2}\sum_{j=2}^{n-1} \beta_j^2\right)$$

c) Prove that (**3 points**)

$$\frac{1}{\sigma^2} \mid \text{data}, \tau, \beta \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\|y - X\beta\|^2\right)$$

d) Implement the Gibbs sampler for the data from Problem 4 (*yahoo* Google trends dataset) and compare the samples obtained from the Gibbs sampler with the samples generated in Problem 4(b). (**9 points**)