Example problem Ch 5

July 22, 2019

Abstract

1 Australian Food Expenditures

The csv file `aus_food.csv` contains records of total monthly expenditure on cafes, restaurants and takeaway food services in Australia ($billion) from 2004 through 2016. (Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on July 22, 2019.)

The file contains two variables: `expenditure` is the monthly expenditure on food services, and `time` encodes the year and month. For example, January 2004 is coded as 2004, and February 2004 is encoded as 2004.0833... since 1/12 is about 0.083. In this example we will build a model to forecast monthly food expenditures. We will fit the model to the data up through 2015, and predict food expenditures in 2016.

1. Read the data in. Make a plot that has food expenditures on the vertical axis and time on the horizontal axis.

2. Your plot in part 1 showed an increasing trend in food expenditures over time, with fairly consistent seasonal patterns (for instance, there is a regular peak in food expenditures in December of each year). Let's use a model for food expenditures that has a linear function of time to capture the long-term increasing trend, and sine and cosine terms to capture seasonality around that trend. Specifically, if we use y_t to denote food expenditures at time t, we have:

$$y_t \approx \alpha_0 + \alpha_1 t + \sum_{k=1}^{K} \{\gamma_k \sin\left(\frac{2\pi k}{12} t\right) + \gamma_k^* \cos\left(\frac{2\pi k}{12} t\right)\}$$

We will use $K = 6$ trigonometric terms. We will explore using sums of trigonometric terms like this more later in the course when we discuss Fourier series. For now, let's just note that the first terms in this sum have $k = 1$, which means those terms have period 12 months, which is reasonable for capturing seasonality with a period of 12 months.

We can express this model in matrix form as $y \approx X\beta$, where for the data from 2004 through 2015, we have:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 2004 & \sin\left(\frac{2\pi}{12}2004\right) & \cos\left(\frac{2\pi}{12}2004\right) & \cdots & \sin\left(\frac{2\pi 6}{12}2004\right) & \cos\left(\frac{2\pi 6}{12}2004\right) \\ 1 & 2004.083 & \sin\left(\frac{2\pi}{12}2004.083\right) & \cos\left(\frac{2\pi}{12}2004.083\right) & \cdots & \sin\left(\frac{2\pi 6}{12}2004.083\right) & \cos\left(\frac{2\pi 6}{12}2004.083\right) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2015.917 & \sin\left(\frac{2\pi}{12}2015.917\right) & \cos\left(\frac{2\pi}{12}2015.917\right) & \cdots & \sin\left(\frac{2\pi 6}{12}2015.917\right) & \cos\left(\frac{2\pi 6}{12}2015.917\right) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \gamma_1 \\ \gamma_1^* \\ \vdots \\ \gamma_6 \\ \gamma_6^* \end{bmatrix}$$

In python, create the matrix X using the data from 2004 up through 2015.

3. Find the least squares estimate of β.

4. Create a matrix X_{new} corresponding to the data for 2016.

5. By combining your estimate of β from part 3 and your X_{new} from part 4, obtain predicted values for food expenditures in every month of 2016.

6. Create a plot comparing your predictions of food expenditure to the actual observed food expenditures in 2016.