STAT 153 & 248 - Time Series
Lecture Thirteen

Spring 2025, UC Berkeley

Aditya Guntuboyina

March 5, 2025

In this lecture, we shall revisit sines and cosines and discuss a high-dimensional model involving sinusoids (there will be connections to the high-dimensional regression models that we studied last week; even though the main model for this week will be somewhat different from those).

1 Recap: Sunspots Data

In order to motivate the model that we shall study today, consider the annual sunspots dataset y_t that we previously looked at multiple times in this class.

Previously (e.g., Lecture 8), we used the following models for the sunspots data:

$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \tag{1}$$

$$y_t = \beta_0 + \beta_1 \cos(2\pi f_1 t) + \beta_2 \sin(2\pi f_1 t) + \beta_3 \cos(2\pi f_2 t) + \beta_4 \sin(2\pi f_2 t) + \epsilon_t \tag{2}$$

$$y_t = \beta_0 + \beta_1 \cos(2\pi f_1 t) + \beta_2 \sin(2\pi f_1 t) + \beta_3 \cos(2\pi f_2 t) + \beta_4 \sin(2\pi f_2 t) + \beta_5 \cos(2\pi f_3 t) + \beta_6 \sin(2\pi f_3 t) + \epsilon_t \tag{3}$$

In all these models, $\epsilon_t \sim N(0, \sigma^2)$. f, f_1, f_2, f_3 represent unknown frequency parameters. These models are helpful for understanding certain aspects of the sunspots data. For example, model (1), when fitted to the data, gave $f \approx 1/11$ which corresponds to the solar cycle. Models (2) and (3) can give reasonable forecasts of the number of sunspots in future years.

In spite of these utilties, these models do not capture many important characteristics of the sunspots dataset. For example, if we generate simulated data $y_1^{\text{simulated}}, \dots, y_n^{\text{simulated}}$ from any of these models (with the parameters β's, f's and σ fixed at the estimates obtained from the sunspots data), these simulated datasets visually look quite different from the actual sunspots data. The sunspots dataset will have well-defined peaks and the gaps between the peaks varies (roughly around 11) from one cycle to another. The simulated datasets will not have such clear peaks.

2 High-dimensional Regression with Sinusoids

One natural attempt to fix issues with the low dimensional models (1), (2) and (3) is to include sinusoid terms for all possible frequencies. There are, in general, infinitely many possible values of frequences (in the range $(0, 0.5]$) but we shall, for simplicity, stick to Fourier frequencies. Recall that, when n is odd, the Fourier frequencies are $1/n, 2/n, \dots, m/n$ where m equals $(n-1)/2$. When n is even, $1/2$ is also a Fourier frequency. We shall stick to the case where n is odd for simplicity.

The high-dimensional analogue of (1), (2), (3) is (note again that n is odd and $m = (n-1)/2$)

$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos(2\pi(j/n)t) + \beta_{2j} \sin(2\pi(j/n)t)\right) + \epsilon_t, \tag{4}$$

where again $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. We can write this model in matrix form as

$$y = X\beta + \epsilon$$

where

$$X = \begin{pmatrix} 1 & \cos(2\pi(1/n)1) & \sin(2\pi(1/n)1) & \cdot & \cdot & \cdot & \cos(2\pi(m/n)1) & \sin(2\pi(m/n)1) \\ 1 & \cos(2\pi(1/n)2) & \sin(2\pi(1/n)2) & \cdot & \cdot & \cdot & \cos(2\pi(m/n)2) & \sin(2\pi(m/n)2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cos(2\pi(1/n)n) & \sin(2\pi(1/n)n) & \cdot & \cdot & \cdot & \cos(2\pi(m/n)n) & \sin(2\pi(m/n)n) \end{pmatrix},$$

and β is the $n \times 1$ vector with components $\beta_0, \beta_{11}, \beta_{21}, \dots, \beta_{1m}, \beta_{2m}$. The number of these coefficient parameters is n so this is a high-dimensional regression model.

If we fit this model via least squares without any regularization by minimizing $\|y - X\beta\|^2$, we would obtain a perfect fit to the data for the choice of parameters:

$$\hat{\beta}_0 = \bar{y} \quad \hat{\beta}_{1k} = \frac{2}{n} \sum_{t=1}^{n} y_t \cos\left(2\pi \frac{k}{n} t\right) \quad \hat{\beta}_{2k} = \frac{2}{n} \sum_{t=1}^{n} y_t \sin\left(2\pi \frac{k}{n} t\right).$$

These can be derived from the orthogonality properties of sinusoids that we discussed previously in Lectures 6 and 7. One can also write $\hat{\beta}_0, \hat{\beta}_{1k}, \hat{\beta}_{2k}, 1 \le k \le m$ in terms of the DFT $b_0, \dots, b_{n-1}$ of y_t.

To prevent overfitting and to obtain something meaningful, we need to add some kind regularization to the high-dimensional model (4). Last week, we look at the ridge and LASSO regularization in high-dimensional linear regression models. In the context of (4), the ridge estimator is given by minimizing

$$\sum_{t=1}^{n} \left(y_t - \beta_0 - \sum_{j=1}^{m} \left(\beta_{1j} \cos(2\pi(j/n)t) + \beta_{2j} \sin(2\pi(j/n)t)\right) \right)^2 + \lambda \sum_{j=1}^{m} \left(\beta_{1j}^2 + \beta_{2j}^2\right). \tag{5}$$

The motivation for this estimator is a desire to obtain estimates of β for which β_{1j}, β_{2j} are somewhat small. In the case of the change of slope model considered last week, smallness of the β-coefficients leads to smooth fits to the data which can be treated as smooth estimates of the underlying trend in the data. However, in the context of the sinusoidal model (4), it is unclear why one would want to obtain small values for β_{1j}, β_{2j}. In fact, for the sunspots data, we expect that β_{1j}, β_{2j} would not be small for some special frequencies (e.g., frequencies j/n which are close to $1/11$). This suggests that the ridge estimator (5) may not yield any anything useful insights when applied to the sunspots dataset.

From our discussion in the last lecture, the ridge estimator (5) can also be understood from the Bayesian perspective under the prior:

$$\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \ldots, \beta_{1m}, \beta_{2m} \overset{\text{i.i.d}}{\sim} N(0, \tau^2). \tag{6}$$

Again, this model makes sense in the change-of-slope ReLU model from last week because it implies that the underlying trend function is smooth but the data, because of somewhat random fluctuations, looks noisy. In the present sinusoidal case, (6) is less justifiable.

3 The Spectrum Model

The spectrum model (described below) is obtained by making two changes to (4) and (6). First the additive error ϵ_t term is dropped from (4). Presence of this term imparts additional random fluctuations to the sinusoidal term, making the model not appropriate for datasets such as the sunspots which appear to not have random fluctuations. This leads to the model:

$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos(2\pi(j/n)t) + \beta_{2j} \sin(2\pi(j/n)t)\right) \tag{7}$$

Secondly, the prior assumption (6) is changed to the following: $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \ldots, \beta_{1m}, \beta_{2m}$ are all independent with

$$\beta_{11}, \beta_{21} \overset{\text{i.i.d}}{\sim} N(0, \tau_1^2), \quad \beta_{12}, \beta_{22} \overset{\text{i.i.d}}{\sim} N(0, \tau_2^2), \quad \ldots \quad , \beta_{1m}, \beta_{2m} \overset{\text{i.i.d}}{\sim} N(0, \tau_m^2).$$

In other words, $\beta_{1j}, \beta_{2j} \overset{\text{i.i.d}}{\sim} N(0, \tau_j^2)$ for $j = 1, \ldots, m$. Instead of using a single τ^2, we now use $\tau_1^2, \ldots, \tau_m^2$ so one variance parameter each for the sinusoid at each frequency j/n.

Taken together the parameters $\tau_1^2, \ldots, \tau_m^2$ are known as the spectrum of the model. The spectrum represents the variances of the random variables that determine the amplitudes of the sinusoidal terms at the Fourier frequencies (see for example page 8 of the book "Spectral analysis for univariate time series" by Percival and Walden).

Under these modeling assumptions, the variance of y_t is given by:

$$\begin{aligned}
\text{var}(y_t) &= \sum_{j=1}^{m} \text{var}\left(\beta_{1j} \cos(2\pi(j/n)t) + \beta_{2j} \sin(2\pi(j/n)t)\right) \\
&= \sum_{j=1}^{m} \left[\text{var}(\beta_{1j}) \cos^2(2\pi(j/n)t) + \text{var}(\beta_{1j}) \sin^2(2\pi(j/n)t)\right] \\
&= \sum_{j=1}^{m} \tau_j^2 \left(\cos^2(2\pi(j/n)t) + \sin^2(2\pi(j/n)t)\right) = \sum_{j=1}^{m} \tau_j^2.
\end{aligned}$$

τ_j^2 represents how much contribution the corresponding frequency j/n has in the overall variance structure of y_t. If τ_j^2 is large for a specific j, the corresponding frequency j/n has a strong contribution to the variance of the data. If τ_j^2 is small, the contribution of that frequency is small.

The sequence $\{\tau_j^2\}$ provides a *spectral representation* of the time series in the sense that it describes the distribution of variance across frequencies.

We shall refer to (7) as the spectrum model. In the next lecture, we shall look at an alternative way of thinking about this model in terms of the DFT of the data.

The spectrum model consists of the unknown parameters β_0 and the spectrum given by the variances $\tau_1^2, \dots, \tau_m^2$. Of these β_0 is not really a parameter as it simply equals $\bar{y}$. To see this, just average both sides of (7) with respect to t and note that

$$\sum_{t=1}^{n} \cos(2\pi(j/n)t) = \sum_{t=1}^{n} \sin(2\pi(j/n)t) = 0$$

which gives $\beta_0 = \bar{y}$. In other words, (7) is equivalent to

$$y_t - \bar{y} = \sum_{j=1}^{m} \left(\beta_{1j} \cos(2\pi(j/n)t) + \beta_{2j} \sin(2\pi(j/n)t)\right) \qquad \text{with } \beta_{1j}, \beta_{2j} \overset{\text{i.i.d}}{\sim} N(0, \tau_j^2).$$

$\tau_1^2, \dots, \tau_m^2$ denote the unknown parameters in this model which will be estimated from the data (we shall see how to do this in the next lecture). Data generated from this model will look differently depending on the exact values of $\tau_1^2, \dots, \tau_m^2$. Here are some examples.

1. Suppose τ_j^2 equals a constant when j/n lies in a fixed interval $[1/9, 1/13]$ and 0 otherwise:

$$\tau_j^2 = \begin{cases} c & \text{if } \frac{j}{n} \in \left[\frac{1}{9}, \frac{1}{13}\right] \\ 0 & \text{otherwise} \end{cases}$$

   Then data generated from this model look periodic with clear peaks. The gaps between the peaks will change from cycle to cycle (some gaps will be 10, some 11, some 9 etc.).

2. Suppose τ_j^2 increases with j. Then the higher frequency sinusoids will dominate, and the data will look quite wiggly.

3. Suppose τ_j^2 decreases with j. Then the lower frequency sinusoids with dominated, giving the data a smoother appearance.