Probabilistic Time Series Analysis: Lab 9

Tim Kunisky

November 14, 2018

A Gaussian Paradox

- For g and h gaussians with $\mathbb{E}g = \mathbb{E}h = 0$, the joint distribution (g, h) depends only on the covariance matrix.
- Consider two pairs (g, h):
  1. g_1 and h_1 are independent $\mathcal{N}(0, 1)$ random variables.
  2. $g_2 \sim \mathcal{N}(0, 1)$, and $h_2 = sg_2$, where $s = \pm 1$ with equal probability and is independent of g_2.

  Both pairs consist of gaussians and have covariance matrix

$$\Sigma = \begin{bmatrix} \mathbb{E}[g_i^2] & \mathbb{E}[g_i h_i] \\ \mathbb{E}[g_i h_i] & \mathbb{E}[h_i^2] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- However, $|g_2| = |h_2|$ always, while $|g_1| \neq |h_1|$ almost surely, so (g_1, h_1) and (g_2, h_2) do not have the same distribution!

What's wrong with this argument?

A Gaussian Process Paradox

- Let's review the basic setup of gaussian process regression: we want to model a time series $f(t)$, so we suppose $f(t)$ is distributed according to a gaussian process with a *parametrized covariance kernel*.

$$\mathbb{E}f(t) = 0,$$
$$\mathbb{E}f(s)f(t) = K(s,t;\boldsymbol{\theta}).$$

- Then, we fit the parameters $\boldsymbol{\theta}$ to the data using (say) maximum likelihood.
- But the model dictates $\mathbb{E}f(t) = 0$, so our model can never capture a trend or seasonality in the data.

What's wrong with this argument?

A Very Special Gaussian Process

- Let's say I do gaussian process regression with a covariance kernel like

$$\mathbb{E} f(s) f(t) = K(s, t; m, c) = m^2 (s - c)(t - c).$$

- The parameters of the kernel are $\boldsymbol{\theta} = (m, c)$.

Consider the covariance matrix of $(f(s_1), \ldots, f(s_k))$ under this model. What can you say about this matrix?

What kinds of functions does this kernel let me model?

Some Computational Aspects

- Let's say I give you a multivariate gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. How would you draw 100,000 samples from it? What would you need to compute? How much time does your method take if $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$?

- Now what if I need to compute the conditional distribution on some $M < N$ of the variables? What's the bottleneck in this calculation?