

Bayesian Generalized Additive Model

Austen

February 13, 2025

1 Generalized Additive Model (GAM)

We have a dataset $D := \{x_i, y_i\}_{i=1}^n$, where x_i is an hour in the day and y_i is the observed number of cars in that hour. As expected, the correlation between the traffic volume and the time of the day is strong, with clear distinct patterns emerging for a different locations/directions and days of the week. To isolate different variables to make conclusions about changes in hourly traffic dynamics, we seek to create a simple model to summarize characteristics of interest. In particular, we are interested in how the rush hour has changed before and after the 2019 pandemic.

One of the most simple but powerful models in data science is linear regression, but as the name suggests isn't suitable for capturing non-linear relationships in data on its own. A common technique to address this constraint is to choose a suitable non-linear mapping such that the mapped inputs can be modeled with linear regression. GAMs are a class of Generalized Linear Models which provides a framework for selecting such a mapping as the sum over *smooth* functions. The smoothness of these functions is desirable for our application because traffic volume as a function of time is "smooth" and we may be interested in rate characteristics of this function which can be obtained through differentiation (since smooth functions are differentiable).

The GAM literature and software packaging calls these smooth function 'smoothers' and many different functions are provided, but we are mostly interested in splines. Spline fitting is similar to fitting polynomials to the data but are generally easier to work with for a variety of reasons (see Runge effect, sklearn example).