Exploratory data analysis

Daeyoung Lim*
Department of Statistics
Korea University

July 11, 2016

1 Pendulum Data

1.1 Data Description

The pendulum data obtained from the paper of David Nott consists of 9 covariates and one target variable. As described in the original paper, it is a simulated data of mechanical pendulum, covariates of which are different parameters of the system and the target variable is the angular velocity.

1.2 Angular Velocity

According to Wikipedia, an angular velocity is *"defined as the rate of change of angular displacement and is a vector quantity (more precisely, a pseudovector) which specifies the angular speed of an object and the axis about which the object is rotating"*. The unit is radians per second.

2 SSGP

For a short summary of Lazaro Gredilla's SSGP, the paper comes up with a decomposition of the function into

$$f(\mathbf{x}) = \sum_{r=1}^{m} a_r \cos\left(2\pi \mathbf{s}_r^\top \mathbf{x}\right) + b_r \sin\left(2\pi \mathbf{s}_r^\top \mathbf{x}\right) \tag{1}$$

where $a_r \sim \mathcal{N}\left(0, m^{-1}\sigma_0^2\right)$, $b_r \sim \mathcal{N}\left(0, m^{-1}\sigma_0^2\right)$. Then, by transforming $\mathbf{x}$ into

$$\phi(\mathbf{x}) = \left[\cos\left(2\pi \mathbf{s}_1^\top \mathbf{x}\right) \quad \sin\left(2\pi \mathbf{s}_1^\top \mathbf{x}\right) \quad \cdots \quad \cos\left(2\pi \mathbf{s}_m^\top \mathbf{x}\right) \quad \sin\left(2\pi \mathbf{s}_m^\top \mathbf{x}\right)\right], \tag{2}$$

the end result becomes similar to linear Gaussian process regression model:

$$\mathrm{E}(\mathbf{y}_*) = \phi(\mathbf{x}_*)^\top A^{-1} \mathbf{\Phi} \mathbf{y} \tag{3}$$

$$\mathrm{Var}(\mathbf{y}_*) = \sigma_n^2 + \sigma_n^2 \phi(\mathbf{x}_*)^\top A^{-1} \phi(\mathbf{x}_*). \tag{4}$$

The author suggests learning the parameters via optimizing the log-marginal likelihood:

$$\log p(\mathbf{y}|\theta) = -\frac{1}{2\sigma_n^2}\left(\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{\Phi}^\top A^{-1} \mathbf{\Phi} \mathbf{y}\right) - \frac{1}{2}\log|A| + m \log \frac{m\sigma_m^2}{\sigma_0^2} - \frac{n}{2}\log\left(2\pi\sigma_n^2\right) \tag{5}$$

by means of the conjugate gradient method.

*Prof. Taeryon Choi

2.1 Sidenote

The predictive mean of Gaussian process regression model is

$$p\left(f_*|\mathbf{x}_*, X, \mathbf{y}\right) = \int p\left(f_*|\mathbf{x}_*, \mathbf{w}\right) p\left(\mathbf{w}|X, \mathbf{y}\right)\, d\mathbf{w} = \int \mathbf{x}_*^\top \mathbf{w} p\left(\mathbf{w}|X, \mathbf{y}\right)\, d\mathbf{w} \tag{6}$$

$$= \mathcal{N}\left(\frac{1}{\sigma_n^2}\mathbf{x}_*^\top A^{-1} X\mathbf{y}, \mathbf{x}_*^\top A^{-1}\mathbf{x}_*\right). \tag{7}$$

The posterior distribution of the weights $\mathbf{w}$ is

$$p\left(\mathbf{w}|X, \mathbf{y}\right) \sim \mathcal{N}\left(\overline{\mathbf{w}} = \frac{1}{\sigma_n^2} A^{-1} X\mathbf{y}, A^{-1}\right) \tag{8}$$

where $A = \sigma_n^{-2} XX^\top + \Sigma_p^{-1}$. Therefore, this is the reason why we can plug in the test data $\widetilde{X}$ into the predictive distribution's $\mathbf{x}_*$ to get the fitted values of the unknown function, $\hat{f}$. (In the case of sparse spectrum decomposition, the design matrix is no longer X but rather Φ.)

3 VA for partially linear additive models

The paper suggests a model of the form

$$y_i = \mu + \sum_{j=1}^{p} f_j\left(x_{ij}\right) + \epsilon_i, \quad i = 1, \ldots, n. \tag{9}$$

- In the authors' exact words, *for simplicity of exposition*, they assumed all covariates to be continuous with support $[0, 1]$.
- Discrete variables are put into the linear part.
- Impose $\mathrm{E}\left(f_j\left(x_j\right)\right) = 0$ constraint to achieve identiability.
- Transform the basis functions in such a way that they are orthogonal to the linear basis functions.
- Essentially VB for parameter estimation but offered some degree of freedom between Monte Carlo estimation and Laplace approximation for the intractable integration terms.

The real data analysis section of the original paper mentions that the crime data in R package {Ecdat} was used, which has 630 observations and 22 variables.

3.1 BPLAM code review

- Since one of the purposes of this paper was to propose a novel way of selecting variables using variational approximation, the code assumes there would be more than one covariate. The code crashes with only `x` of one column.
- The code still worked even though one of the covariates was outside of the range $[0, 1]$.

4 Kneib VA

Model:

$$y_i = \mathbf{x}^\top \boldsymbol{\beta} + f_1\left(\text{herdsize}_i\right) + f_2\left(\text{capital}_i\right) + f_{\text{spat}}\left(\text{county}_i\right) + \epsilon_i. \tag{10}$$

Code doesn't work...

5 To be done

- Fix Kneib's code
- EDA Miguel's datasets:
  - `elevators` data
  - `pumadyn32nm` data
  - `pendulum` data
  - `kin40k` data
  - `pol` data

  and BPLAM real data set: Ecdat crime data.
- Run BSAR VB/MCMC (no restriction/shape restriction), SSGP, spline VB, Kneib VB
  - real data sets (listed above)
  - simulated data: 3 models in Kneib's paper, 1 model in BPLAM etc

  and compare RMSE, code execution time etc.

6 Datasets

6.1 Pole Telecomm and Elevators

According to Lazaro Gredilla, the data sets are taken from `http://www.liaad.up.pt/ ltorgo/Regression/Dat`

7 Transforming data with support $\mathcal{S}$

Normally, the data don't lie within the interval $[0,1]$. Therefore, we must transform them according to their support. One of the distributions with universal support is the Cauchy distribution. The pdf and cdf are

$$q(x) = \frac{1}{\pi(1+x^2)} \tag{11}$$

$$Q(x) = \frac{1}{\pi}\tan^{-1}(x) + \frac{1}{2}. \tag{12}$$

Now define $\varphi_0(x) = \sqrt{q(x)}$ and $\varphi_j(x) = \sqrt{2q(x)}\cos[\pi j Q(x)]$. Let's do one of the shape restriction models. Recall

$$\varphi^a_{j,k}(x) = \int_0^x \varphi_j(s)\,\overline{\varphi}_k(s)\,ds - \int_0^1\int_0^s \varphi_j(t)\,\overline{\varphi}_k(t)\,dt\,ds \quad \text{for } j,k \geq 0. \tag{13}$$

With the defined basis functions above,

$$\int_0^x \varphi_j(s)\,\overline{\varphi}_k(s)\,dx = \sqrt{2}\int_0^x \frac{1}{\pi(1+x^2)}\cos\left[j\tan^{-1}(s)+\frac{\pi j}{2}\right]ds \tag{14}$$

$$= \sqrt{2}\int_{Q(0)}^{Q(x)} \cos[\pi j u]\,du \quad (u = Q(x)) \tag{15}$$

$$= \frac{\sqrt{2}}{\pi j}\left\{\sin\left(j\tan^{-1}(x)+\frac{\pi j}{2}\right)-\sin\left(\frac{\pi j}{2}\right)\right\} \tag{16}$$

$$\int_0^1\int_0^s \varphi_j(t)\,\overline{\varphi}_k(t)\,dt\,ds = \int_0^1\int_{Q(0)}^{Q(s)} 2\cos[\pi j u]\cos[\pi k u]\,du\,ds \tag{17}$$

$$= \int_0^1 \frac{\sin(\pi(j-k)Q(s))}{\pi(j-k)}+\frac{\sin(\pi(j+k)Q(s))}{\pi(j+k)}\,ds \tag{18}$$

$$-\frac{\sin(\pi(j-k)/2)}{\pi(j-k)}-\frac{\sin(\pi(j+k)/2)}{\pi(j+k)} \tag{19}$$

Comparing with the one with support $[0,1]$ suggested in the original paper,

$$\int_0^x \varphi_j(s)\,\overline{\varphi}_k(s)\,ds = \frac{\sqrt{2}}{\pi j}\sin(\pi j x) \tag{20}$$

there is another term added and care must be taken in that we shouldn't simply apply the cdf to the data and do the rest equally. The math tells us it becomes different.