Learning Image with Gaussian Process Regression and Application to Classification

UNIVERSITY OF SOUTH CAROLINA

Tahmidul Islam

Department of Statistics, University of South Carolina

Goal

Image classification is a common machine learning task with numerous tools available. We explore whether if we can learn a probability distribution generating the images and further use it for classification. We can summarize our goal as the following.

- Learn the joint distribution of the pixel intensities using Gaussian process (GP) regression by assuming spatial correlation between the pixels.
- Develop an efficient GP regression training algorithm for replicated responses on few unique covariate points. This special structure is present in the image data.
- Use this joint distribution in a Bayes classifier to classify images. We relax the conditional independence assumption of naive Bayes classifier and investigate improvement in the classification accuracy.

GP Regression

Gaussian process (GP) is a stochastic process $\{f(x) : x \in \mathcal{X}\}$. In Bayesian nonparametric regression setting, GP with kernel $K(\cdot,\cdot)$ is used as a prior on function as:

$$
\begin{aligned}
Y &= f(x) + \epsilon \\
f(\cdot) &\sim \mathcal{GP}(\mu(\cdot), K(\cdot,\cdot)) \\
\epsilon &\sim N(0, \sigma^2).
\end{aligned}
$$

The posterior distribution can be made tractable by assuming normally distributed error term. The posterior predictive distribution for test points X^* is a multivariate normal distribution with following mean and covariance matrix.

$$
\begin{aligned}
\boldsymbol{\mu}^* &= K(X^*, X)(K(X,X) + \sigma^2 I)^{-1}\boldsymbol{y} \\
\Sigma^* &= K*(X^*, X^*) + \sigma^2 I - \\
&\quad K(X^*, X)(K(X,X) + \sigma^2 I)^{-1} K(X, X^*).
\end{aligned}
$$

GP Regression Model for Image

We adopt a functional data type model. Let $Y_{ij}(x,z)$ be the pixel intensities of the j^{th} image in the i^{th} class. The pixel intensities are assumed to be a fixed function $f(\cdot)$ of the spatial location of the pixels. A random error function introduces variation between the images. A white noise measurement error is considered.

$$
\begin{aligned}
Y_{ij}(x,z) &= f(x,z) + g_{ij}(x,z) + \epsilon_{ij}(x,z) \\
f(\cdot) &\sim \mathcal{GP}(0, K_i) \\
g_{ij}(\cdot) &\sim \mathcal{GP}(0, W_i) \\
\epsilon_{ij}(x,z) &\sim N(0, \sigma^2).
\end{aligned}
$$

For (28×28) images, only 1000 training images will produce a training data of size $784,000$. We develop a computationally efficient algorithm to fit such data by considering a special case of the covariance structure where W_i is taken to be diagonal covariance kernel. We write a ANOVA type model since the observations are coming from p^2 unique location.

$$
\mathbf{Y} = \mathbf{A}\mathbf{f} + \boldsymbol{\epsilon}.
$$

$$
\mathbf{A} = \begin{bmatrix} \mathbb{1}_n & & & \\ & \mathbb{1}_n & & \\ & & \ddots & \\ & & & \mathbb{1}_n \end{bmatrix}_{np^2 \times p^2},
\mathbf{f} = \begin{bmatrix} f(1,1) \\ f(1,2) \\ \vdots \\ f(p,p) \end{bmatrix}_{p^2 \times 1},
\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{1,1} \\ \epsilon_{1,2} \\ \vdots \\ \epsilon_{p,p} \end{bmatrix}_{np^2 \times 1},
\mathbb{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}.
$$

This representation leads to reduction in $\mathcal{O}(n^3 p^6)$ computation complexity to $\mathcal{O}(p^6)$ only. This algorithm is also applicable when performing general GP regression with replications. If there is N total training data but only N^* unique covariate values then computational complexity will be $\mathcal{O}(N^{*3})$ rather than $\mathcal{O}(N^3)$. We compute the posterior mean and the covariance matrix which is further used for classification.

Learning Image and Classification

We obtain the posterior distribution of the pixels on the same spacial locations separately for each class which considered as the joint distribution of the pixel intensities, i.e. the image. We applied this method on the MNIST dataset.

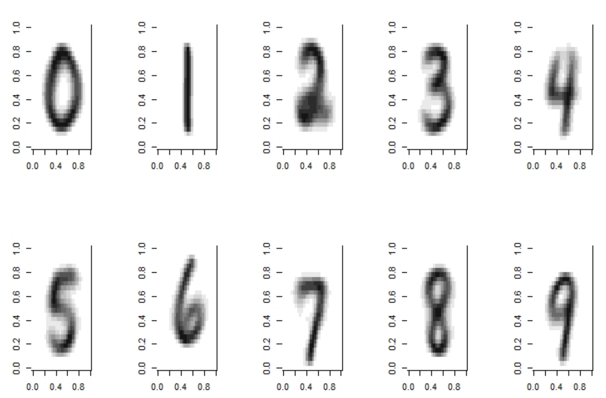

Figure 1: Posterior means (Gaussian Kernel)

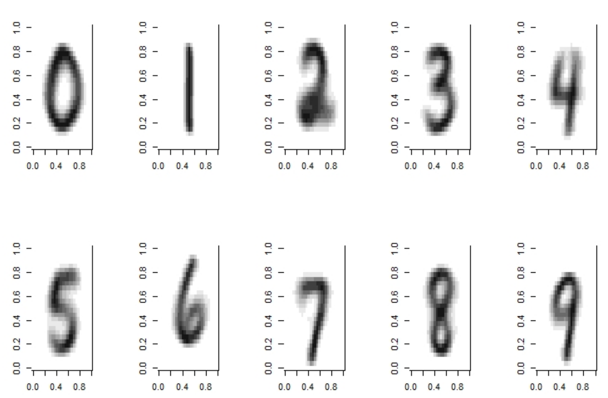

Figure 2: Posterior means (Laplace Kernel)

This posterior enables us to use the Bayes classifier.

$$
H(\tilde{\mathbf{y}}) = \underset{c \in \{1,2,\ldots,m\}}{argmax}\; P_{\mathbf{Y}^*}(\tilde{\mathbf{y}} | \mathbf{y_1}, \mathbf{y_2}, \ldots \mathbf{y_n}, C = c)\hat{P}(C = c).
$$

Discussion

- We show how we can estimate the joint distribution of the pixel intensities using GP regression which requires small handful of hyperparameter estimation.
- A computationally feasible method for fitting GP regression for replicated data is developed.
- Posterior is used to classify images in a Bayes classifier framework. We applied this method on (MNIST) dataset. It provides about 87% accuracy on the test set. Where the naive Bayes classifier achieves about 53% accuracy. This shows that replacing the conditional independence assumption by spatially correlated pixels boosts performance.
- Stationary kernels (Gaussian or Laplace) were tried. Use of non-stationary kernels and more flexible error assumption might improve the results.
- Although only image data was considered, this is a general method for functional data in any dimension. We are working on more generalized framework of classifying irregular/regular functional data.

Reference

[1] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[3] Gareth M James and Trevor J Hastie. Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):533–550, 2001.