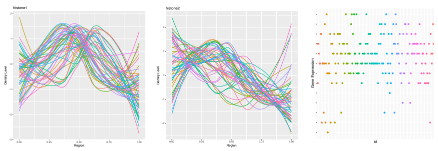
University of Georgia, Big Data Analytics Lab

Optimal Generalized Quadrature Functional Regression in Reproducing Kernel Hilbert Space

Honghe Jin[1], Ping Ma[1], Wenxuan Zhong[1]

1 Department of Statistics, University of Georgia. 310 Herty Drive, Athens, GA.

Background: Generalized Functional Regression



Functional predictor | Functional predictor | Poisson response

GLM can be generalized to the case that the covariates are functions. We estimate the coefficient function β in a reproducing kernel Hilbert space in the following model.

- **Model setup:** Let $\{X_i(r,s) \in \mathcal{L}_2[I_s \times I_r], i = 1, \dots, n, s \in I_s, r \in I_r\}$ be random processes. Given X_i, the response follows an exponential family distribution

$$f_Y(y|X) = \exp\left\{\frac{y\eta(X) - b(\eta(X))}{a(\phi)}\right\} + c(y, \phi),$$

where $a > 0, b$, and c are known functions, $\boldsymbol{\eta}(\boldsymbol{X}) = \int_{I_s}\int_{I_r} \boldsymbol{X}(\boldsymbol{r},\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{r},\boldsymbol{s})\, \boldsymbol{drds}$ is the canonical parameter with the parameter function **$\boldsymbol{\beta}$ to be estimated**, and ϕ is either known or a nuisance parameter.

- **Alternative forms:**

$$\eta(X,Z) = \beta_0 + \int_{I_s}\int_{I_r} X(r)\beta(r,s)Z(s)\,drds$$
$$\eta(X,Z) = \int_{I_s} X(r)\gamma(r)dr + \int_{I_s}\int_{I_r} X(r)\beta(r,s)Z(s)\,drds$$

……

Motivating Examples

- **Logistic Regression.** Consider binary responses $Y_i|X_i \sim \text{Bin}(p_i)$, where $p_i = \frac{\exp(\eta(X_i))}{1+\exp(\eta(X_i))}$ and the density is

$$f(y|x) = p^y(1-p)^{1-y} = \exp(y\eta(x) - \log(1 + e^{\eta(x)}),$$

where $\eta(x) = \int_{I_s}\int_{I_r} x(r,s)\beta(r,s)drds = \log(\frac{p(x)}{1-p(x)})$ is the logit function. In this case, $b(\eta) = \log(1 + e^{\eta})$ and $a(\phi) = 1$.

Motivation: Image classification.

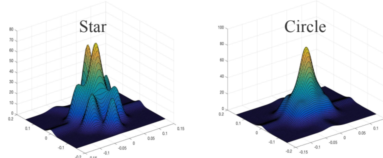


Figure 1: Density functions of gradients of a star and a circle.

- **Poisson regression:** Consider Poisson responses $Y_i|X_i \sim \text{Poi}(\lambda_i)$, where $\lambda_i = \exp(\eta(X_i))$, and the density is

$$f(y|x) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\{y\eta(x) - e^{\eta(x)} - \log(y!)\}$$

where $\eta(x) = \int_{I_s}\int_{I_r} x(r,s)\beta(r,s)drds = \log(\lambda(x))$ is the log intensity. In this case, $b(\eta) = e^{\eta}$ and $a(\phi) = 1$.

Motivation: Gene expression effected by Histone Modification

Y: Gene expression value
$X(r)$: Histone modification level *H3K9me3*
$Z(s)$: Histone modification level *H3K4me2*

Penalized Likelihood Functional Regression

To estimate the coefficient function β, one can minimize the following penalized likelihood functional

$$-\frac{1}{n}\sum_{i=1}^{n}\{Y_i\eta(X_i) - b(\eta(X_i))\} + \lambda J(\beta) \qquad (1)$$

where $J(\beta)$ is a quadratic functional to quantify the smoothness of β, and λ is the smoothing parameter balancing the tradeoff between the goodness of fit and the smoothness of β.

Suppose β is located in a reproducing kernel Hilbert space $\mathcal{H}$ with the reproducing kernel K, then the minimizer of (1) can be reduced into **a finite-dimensional space**.

Theorem 1. *Denote $\hat\beta$ as the minimizer of (1), then $\hat\beta$ has the form*

$$\hat\beta = [\sum_{j=1}^{N_r} d_j^r\psi_j^r(r) + \sum_{i=1}^{n} c_i^r\xi_i^r(r)][\sum_{j=1}^{N_s} d_j^s\psi_j^s(s) + \sum_{i=1}^{n} c_i^s\xi_i^s(s)]$$
$$= \boldsymbol{d}^T\boldsymbol{\psi}(r,s) + \boldsymbol{c}^T\boldsymbol{\xi}(r,s)$$

where d^r, c^r, d^s, c^s are coefficients to estimate, $\psi^r, \xi^r, \psi^s, \xi^s$ are basis functions of the corresponding marginal reproducing kernel Hilbert spaces.

Estimation of the Coefficient Function

It is easy to check that (1) is strictly convex in η, and hence strictly convex in β. Therefore, one can perform Newton iteration to calculate the minimizer $\hat\beta$ given a fixing smoothing parameter λ. Similar to the classic GLM, the estimation of β can be updated by minimizing the penalized weighted least squares

$$\frac{1}{n}\sum_{i=1}^{n} w_i\left(\tilde{Y}_i - \eta(X_i,\beta)\right)^2 + \lambda J(\beta) \qquad (2)$$

The smoothing parameter λ can be selected by minimizing the generalized approximate cross validation (GACV).

Let S be an $n \times N$ matrix with the $(i,j)^{th}$ entry $\int_{I_s}\int_{I_r} X_i(r,s)\psi_j(r,s)drds$, R be an $n \times M$ matrix with the $(i,k)^{th}$ entry $\int_{I_s}\int_{I_r} X_i(r,s)\xi_j(r,s)drds$, and Σ be an $M \times M$ matrix with the $(i,j)^{\text{th}}$ entry $\langle \xi_i, \xi_j \rangle = J(\xi_i, \xi_j)$. $N = N_r N_s$, $M = n^2 + n(N_r + N_s)$. Then estimating β in (2) is reduced to find $\boldsymbol{d}$ and $\boldsymbol{c}$ in

Quadrature form equivalent to the objective function

$$\min_{c,d} (Y_w - S_w\boldsymbol{d} - R_w\boldsymbol{c})^T(Y_w - S_w\boldsymbol{d} - R_w\boldsymbol{c}) + n\lambda\boldsymbol{c}^T\Sigma\boldsymbol{c}$$

where $Y_w = W^{1/2}Y, S_w = W^{1/2}S, R_w = W^{1/2}R, W = \text{diag}(w_1, \dots, w_n)$.

Optimal Convergence Rate

We focus on the prediction error $E(\hat\eta - \eta)^2$, which can be bounded by

$$\mathcal{R}_n \triangleq \int\int\int\int \left(\beta(r_1,s_1) - \hat\beta(r_1,s_1)\right) M(r_1,s_1;r_2,s_2)\left(\beta(r_2,s_2) - \hat\beta(r_2,s_2)\right)dr_1ds_1dr_2ds_2$$

where M is the covariance kernel of X, $\mathcal{R}_n$ is the risk based on the data with the sample size n.

Some conditions are needed to develop the convergence rate

- *The eigenvalues ρ_ν of the kernel $K^{1/2}MK^{1/2}$ is of the order $\rho_\nu \asymp \nu^{-2r}$.*
- *The kurtosis of $\eta(X,\beta)$ is bounded for any $\beta \in \mathcal{L}_2$*
- *The b' is monotonic, b'' and $b^{(3)}$ are uniformly bounded*

Theorem 2. *If the conditions above hold, and $\hat\beta$ minimizes* (1), *then*

$$\lim_{n\to\infty}\sup_{\beta\in\mathcal{H}} \mathcal{R}_n = O_p(\lambda + n^{-1}\lambda^{-1/2r} + n^{-1}) \qquad (3)$$

In addition, we have the minimax convergence rate over all possible estimators based on the data with sample size n

Theorem 3. *If the conditions above hold, then*

$$\lim_{n\to\infty}\inf_{\tilde\beta}\sup_{\beta\in\mathcal{H}} \mathbb{P}(\mathcal{R}_n \geq cn^{-\frac{2r}{1+2r}}) = 1,$$

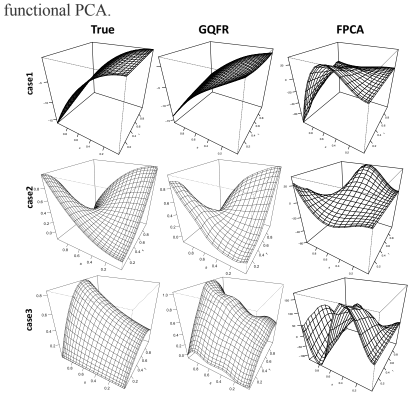
for any $c \to \infty$, where $\tilde\beta$ are taken over all possible estimators given the training samples and $\mathcal{R}_n$ is corresponding to $\tilde\beta$.

Taking $\lambda = O(n^{-\frac{2r}{1+2r}})$ in (3), then (3) achieves the best convergence rate at $O(n^{-\frac{2r}{1+2r}})$.

Combining **Theorems 2** and **Theorem 3**, we have shown the convergence rate of our estimator achieves the minimax lower bound, and hence our estimator is a **rate optimal estimator**.

Simulation Results

We generate 500 samples to do the functional logistic regression in three cases and compare the proposed method (GQFR) with functional PCA.



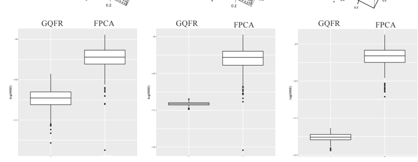
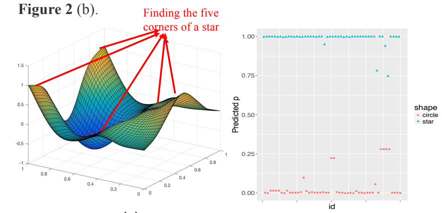


Image Classification

We use the density functions, i.e., **Figure 1**, of 50 stars and 50 circles from ImageNet to train the model. The estimated coefficient function is shown in **Figure 2** (a), and the fitted probability to be a star in **Figure 2** (b).



(a) **Figure 2** (b)

We compare our method with FPCA, as well as SVM to classify the images with 500 testing images.

Training size	50	100	500
SVM	69%	81%	97%
FPCA	63%	72%	71%
GQFR	**93%**	**99%**	**100%**

Histone Modification

A histone modification is a vital post-translational modification to histone proteins. Quantitative analysis of the correlation between histone modifications and gene expression is crucial for the development of histone modifying enzyme-targeted drugs and the better understanding of epigenetic regulation of cellular processes.

We study the regulation mechanism between gene expression and two types of histone modifications, i.e. *H3K9me3* and *H3K4me2*, for liver cancer cell line, HepG2.

	R^2	CV
FPCA	0.093	7.441
GQFR	**0.322**	**3.909**

Discussion

In this project, a penalized generalized quadrature functional regression is proposed to estimate the coefficient function in a reproducing kernel Hilbert space.

- The estimator is a projection of the true function in a finite-dimensional space;
- The estimator achieves the optimal convergence rate;
- The proposed method is applied on two real data examples.

References

[1] Gu, C. (2013). Smoothing spline ANOVA models, Volume 297. Springer Science & Business Media.
[2] Du, P. and X. Wang (2014). Penalized likelihood functional regression. Statistica Sinica, 1017–1041.
[3] Yuan, M., T. T. Cai, et al. (2010). A reproducing kernel Hilbert space approach to functional linear regression. The Annals of Statistics 38 (6), 3412–3444.
[4] Muller, H.-G. and Stadtmulller, U. (2005). Generalized functional linear models. The Annals of Statistics 33, 774-805.

Big Data Analytics Lab