Nonlinear Dimension Reduction for Conditional Quantiles

Dr. Eliana Christou, Annabel Settle, Dr. Andreas Artemiou

Department of Mathematics and Statistics, University of North Carolina at Charlotte; School of Mathematics, Cardiff University

Introduction

• Quantile regression has received a considerable amount of attention as a means of capturing a more complete picture of the conditional distribution of a response and the predictors. However, this extra information can become a hindrance when working with high dimensional data, making dimension reduction techniques particularly useful.

The ability to condense the number of predictors needed–while still retaining all relevant information of the relationship between the independent and dependent variables–allows for increased accuracy, efficiency and decreased computational costs.

Central Quantile Subspace

• In one of the most frequently cited papers on dimension reduction, Li [4] coined the central subspace (CS). The CS is the smallest dimension reduction subspace, defined as the column space of any matrix $\mathbf{A}$ such that $Y \perp\!\!\!\perp \mathbf{X}|\mathbf{A}^T\mathbf{X}$. However, this describes the entire conditional distribution. When only certain relationships are of interest, the CS provides more directions than necessary.

• Christou [3] introduces the τ-th central quantile subspace (CQS). For $\tau \in [0, 1]$ and any matrix $\mathbf{B}\tau$, a τ-th central quantile subspace is the space spanned by the columns of

$\mathbf{B}\tau$ such that $Y \perp\!\!\!\perp Q_\tau(Y|\mathbf{X})|\mathbf{B}\tau^T X$ (1)

• To describe the intersection of all the τ-th quantile dimension reduction subspaces, the CQS provides only the pertinent directions allowing for better dimension reduction. To estimate the CQS, a non-iterative linear algorithm is developed. However, by focusing on only the linear combinations of the predictor matrix $\mathbf{X}$, many important nonlinear features of the data are overlooked.

Therefore, a nonlinear extension of this algorithm is desirable.

The Kernel Method

• To construct the nonlinear CQS, we utilize the "kernel-trick". When trying to separate data that is not linearly separable in the input space, in some instances transforming it into a higher dimensional feature space transforms the data cloud, permitting linear separation. This is the essence of the "kernel-trick" and it's important implication: applying a linear algorithm in the feature space corresponds to a nonlinear algorithm in the original space.

• To generalize the algorithm, the linear estimator of (1) $\mathbf{B}\tau^T(\mathbf{X})$ is replaced with the nonlinear $\Psi_\tau(\mathbf{X})$ under the same assumptions. Following the kernel sliced inverse regression methods of [5] and [2], we map the input space $\chi \subset \mathbb{R}^p$ to the reproducing kernel Hilbert space generated by a positive-definite kernel K, $\mathcal{H}_K$

• Since the feature space K($\mathbf{X}$,·) and the feature τ-CQS directions are in high dimensional space, a finite basis is used to estimate the parameters, the process is discussed in further detail in [1]. At the sample level, let $\{Y_i, \mathbf{X}_i\}_{i=1}^n$ be iid observations. Once the kernel matrix $\{K(\mathbf{X}_i, \mathbf{X}_j)\}_{i,j=1}^n$ and the new predictors $\mathbf{T}_i = (K(\mathbf{X}_i, \mathbf{X}_1), ..., K(\mathbf{X}_i, \mathbf{X}_n))^T$ have been formed, we apply the algorithm of [3] to the data $\{Y_i, \mathbf{T}_i\}_{i=1}^n$ to obtain an estimated basis matrix for the feature τ-CQS. From this basis matrix we can form the new feature τ-CQS predictors and use existing QR techniques to estimate the conditional quantile function.

Data Visualization

To illustrate the performance of the feature τ-CQS, denoted by kernel CQS (KCQS) in the plots, we apply the methodology to the following real world data sets:

• *Ionosphere* containing information on radar returns collected by Johns Hopkins University. The dependent variable of interests labels "good" or "bad" radar returns, with the other 34 variables describing the discrete values of the real and imaginary parts of an ACF.

• The *Heart* dataset contains 76 attributes, though traditionally a subset of 14 is used, predicting the presence of heart disease in a patient. The response is integer valued 0 to 4 where 0 is the absence of heart disease.

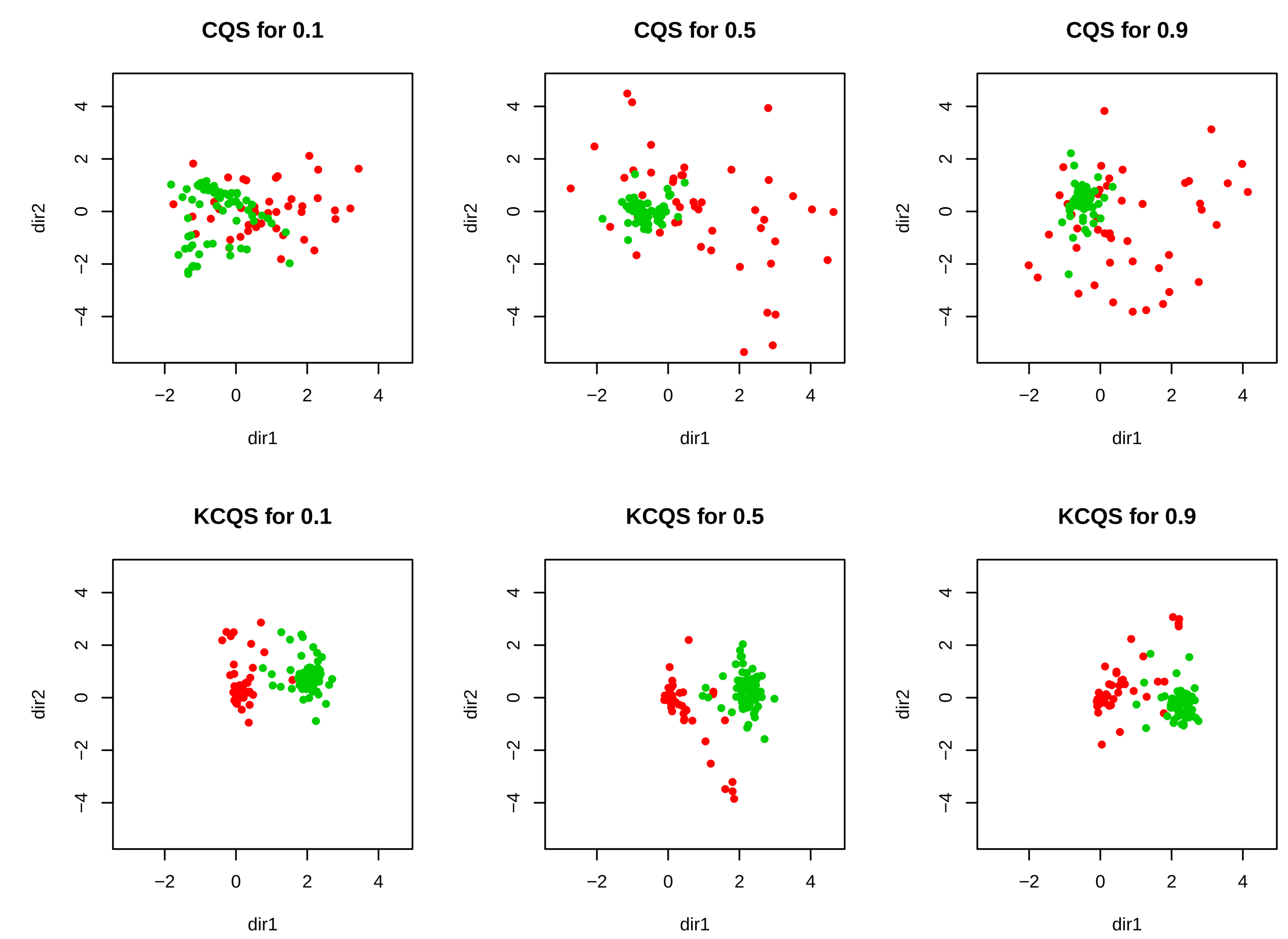

Fig. 1: Visualization of Ionosphere data set

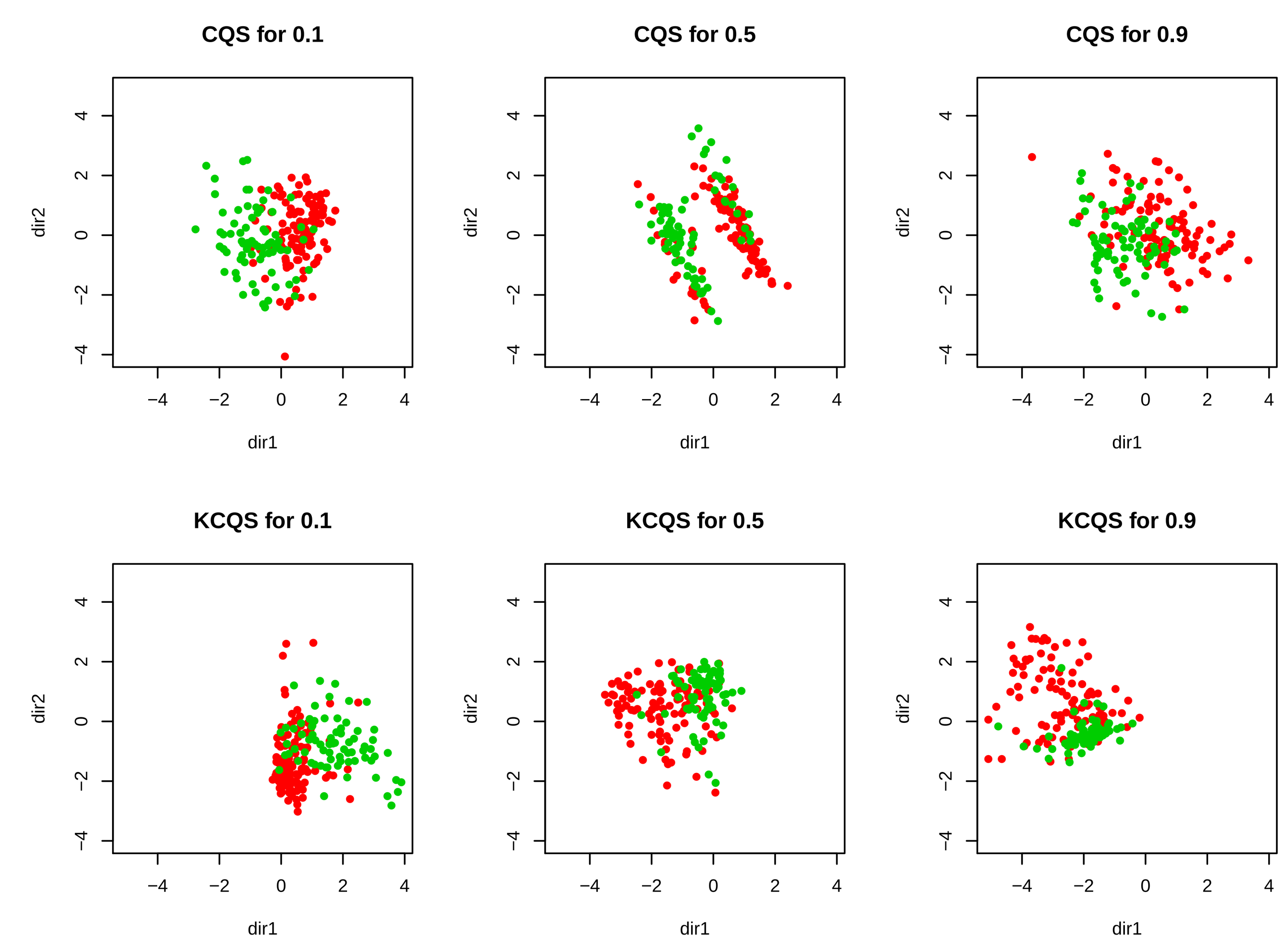

Fig. 2: Visualization of Heart data set

Results

• To demonstrate the ability of a low-dimensional subspace in capturing the main data structure, we use a training set to extract the linear and feature τ-CQS sufficient dimension reduction directions. Then, the first two directions are plotted using a test set, where the data is split into 1/3 and 2/3 subsets for training and testing, respectively.

• Figure 1 showcases the capabilities of the methodology for data where the linear algorithm in the input space performed satisfactory. However, there is clearly a better defined pattern for the K-CQS.

• From Figure 2, we can see the nonlinear algorithm performs well even when there is a significant amount of overlap between classes.

• We can see the reduced nonlinear variables derived from the feature τ-CQS are better able to distinctly separate the data in the two-dimensional subspace compared to the τ-CQS of [3].

Conclusion

• Building on the work of [3], we extend the linear algorithm for estimating the τ-CQS utilizing the "kernel-trick", wherein applying the linear algorithm in the feature space corresponds to a nonlinear algorithm in the input space.

• Real data application, specifically data visualization, highlight the ability of the feature τ-CQS to extract the sufficient dimension reduction directions and capture the data in a low-dimensional space.

• Similarly, the feature τ-CQS performance demonstrates the ability of the methodology and nonlinear dimension reduction techniques to better describe the conditional quantiles of modern data.

References

[1] E. Christou et al. "Nonlinear Dimension Reduction for Conditional Quantiles". In: *Manuscript submitted for publication* ().

[2] Y.-R. Yeh et al. "Nonlinear dimension reduction with kernel sliced inverse regression". In: *IEEE Transactions on Knowledge and Data Engineering* 21 (2009), pp. 1590–1603.

[3] E. Christou. "Central Quantile Subspace". In: *Statistics and Computing* 30 (2020), pp. 677–695.

[4] K.-C. Li. "Sliced Inverse Regression for Dimension Reduction". In: *Journal of the American Statistical Association* 86 (1991), pp. 316–327.

[5] H.-M. Wu. "Kernel sliced inverse regression with applications to classification". In: *Journal of Computational and Graphical Statistics* 17 (2008), pp. 590–610.