NIH Public Access
Author Manuscript
Proc Workshop Math Methods Biomed Image Analysis. Author manuscript; available in PMC 2013 October 25.

Published in final edited form as:
Proc Workshop Math Methods Biomed Image Analysis. 2012 January 10; 2012: . doi:10.1109/MMBIA.2012.6164735.

Max Margin General Linear Modeling for Neuroimage Analyses

Nagesh Adluru[*], **Chad M. Ennis**, **Richard J. Davidson**, and **Andrew L. Alexander**
University of Wisconsin-Madison

Nagesh Adluru: adluru@wisc.edu

Abstract

General linear modeling (GLM) is one of the most commonly used approaches to perform voxel based analyses (VBA) for hypotheses testing in neuroimaging. In this paper we tie support vector machine based regression (SVR) and classical significance testing to provide the benefits of max margin estimation in the GLM setting. Using Welch-Satterthwaite approximations, we compute degrees of freedom (df) of error (also known as residual df) for SVR. We demonstrate that SVR can result not only in robustness of estimation but also improved residual df compared to the very commonly used ordinary least squares (OLS) estimation. This can result in higher sensitivity to signal in neuroimaging studies and also allow for better control of confounding effects of nuisance covariates. We demonstrate the application of our approach in white matter analyses using diffusion tensor imaging (DTI) data from autism and emotion-regulation studies.

1. Voxel Based Analyses

Voxel based analyses (VBA) are typically used to identify imaging phenotypes of a disease group. Support vector machines have been used in neuroimage analyses mostly in the context of classification but not in the context of GLM. Below we elucidate top-level modeling differences between the two exercises. In both cases, we have two types of data for n different subjects: (1) brain data, (2) behavioral/physiological/diagnostic data.

1.1. Data Modeling

VBA is typically based on generalized linear modeling (GLM). GLM is based on the assumption that the brain signal can be explained by a linear combination of a set of design (explanatory) variables. The signal can be either univariate or multivariate (vector-valued) [18]. The elegance of GLM is in the unification of various statistical inferences, like analysis of variance (ANOVA) and covariance (ANCOVA), into regression analyses.

Let us assume there are v voxels in the brain, then VBA works on the following modeling at each voxel:

$$Y = X\beta + \varepsilon, \text{ where } Y \in \mathbb{R}^{n\times 1}, X \in \mathbb{R}^{n\times(p+1)}, \beta \in \mathbb{R}^{(p+1)\times 1},$$

where Y is the observed signal, X is the design matrix of observed p(n) explanatory variables and a column of constants. is a vector indicating the effect of each variable on the signal and also the intercept.

In contrast the data in classification is modeled as:

[*]Research partially supported by NIH P50-MH84051, NIH RO1 MH08026, NIH P30 HD003352, and the Henry M. Jackson Foundation.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$Y = X\beta, \text{ where } Y \in \{-1, 1\}^{n\times 1}, X \in \mathbb{R}^{n\times v}, \beta \in \mathbb{R}^{v\times 1},$$

where now, X is a matrix of vectorized brain signal also known as the feature matrix, Y is the diagnostic information. Each brain is treated as a high-dimensional feature vector. Hence, the key difference in β is that it is a very high-dimensional object ($v \gg n$), in classification but it is *not* a high-dimensional object ($p \ll n$) in the GLM setting.

1.2. Model Estimation

Below we present the difference between OLS and SVR in the context of GLM estimation. Ordinary least squares (OLS) is one of the most commonly used approaches in VBA today, which estimates β by minimizing least squares of the residuals:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} ||Y - X\beta||^2, \quad (1)$$

which gives to a closed form solution: $\beta = (X^T X)^{-1} X^T Y$.

In ε-SVR, the goal becomes to minimize the following objective function:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} ||\beta||^2, \quad s.t. \forall i \in \{1, 2, \ldots, n\}, |y_i - (x_i\beta)| \leq \varepsilon, \quad (2)$$

where y_i and x_is are i^{th} rows in Y and X respectively. In practice, to account for feasibility of constraints the following relaxed version with slack variables for the constraints is solved numerically. This also allows for a trade-off between the regularizer on β as well as the errors in the constraints by controlling C [16].

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(||\beta||^2 + C\sum_{i=1}^{n} (\xi_i + \xi_i^*) \right), \quad s.t. \forall i \begin{cases} y_i - x_i\beta & \leq \varepsilon + \xi_i; \\ x_i\beta - y_i & \leq \varepsilon + \xi_i^* \end{cases}. \quad (3)$$

The main difference between OLS and SVR is that the former has implicit regularization on β by using $\|Y - X\beta\|^2$, while SVR has $|Y - X\beta|_\varepsilon$ as the loss function with $\|\beta\|^2$ as the additional regularization. The ε-insensitivity and the L_1-loss give robustness to the estimation, since it gives more weight to smaller residuals compared to least squares of OLS, which gives most weight to the largest residuals [8]. This estimation procedure gets various names: in the primal form (Eq. (3)) the regularization term on β gives its name "max-margin machine" since in the classification setting $\frac{2}{||\beta||}$ equals the margin of separation between two classes. The dual form of Eq. (3) gives rise to names like "support vector" and "kernel" machines. Since our focus is on using linear kernels and measuring significance of the model (βs) rather than just the accuracy of the predicted output values ($\hat{Y}$), we propose to use the name max-margin GLM.

1.3. Main Contributions

In this paper, motivated by its success in machine learning applications, we propose to use max-margin estimation for GLM. This estimation provides not only robustness to outliers but also provides improved *residual* degrees of freedom (rdf) thus enabling higher sensitivity to signal in neuroimage analyses. It also allows for better accounting of variance of the nuisance covariates in the regression model. Our key mathematical contribution is in computing the rdf of ε-SVR using the Welch-Satterthwaite approximation [12]. This

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

allowed us to integrate the SVR into GLM for improved statistical inference on models. Furthermore, since we perform such an integration in widely used software packages, we can hope for a more direct impact on neuroimaging analyses. We would also like to highlight the difference between degrees of freedom of the model (dfm) and rdf. Gunter and Zhu [5] and the references within worked on computing the effective degrees of freedom of the SVR, but not the rdf. To our best knowledge rdf of SVR has not been computed before and certainly not in the context of GLM for neuroimage analyses.

2. Hypotheses Testing

In VBA, generally one wants to test if a linear combination of the βs is statistically significant. That is, at each voxel, the null-hypothesis tested is: H_0: $\tau\beta = 0$, where τ is an $m \times p$ matrix typically called a *contrast* matrix and the alternative hypothesis is H_1: $\tau\beta \neq 0$. For example, consider the following GLM:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon, \quad (4)$$

where X_i is the i^{th} column of the design matrix. If one wants to test the null-hypotheses, $\beta_1 = 2\beta_2$ and $\beta_3 = \beta_4$ then:

$$\mathcal{T} = \begin{bmatrix} 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}. \quad (5)$$

This method of using contrast matrices provides a general way of representing null-hypotheses. In order to reject null hypotheses we need to compute: $p(\tau\beta \neq 0 | H_0)$ which is the probability of choosing alternate hypotheses when null-hypotheses are true. In other words it is the probability of false rejection of null-hypotheses or false discovery of alternate hypotheses. It is also typically called *p*-value. If this value is smaller than a certain threshold α then one can reject the null hypotheses with $1 - \alpha$ confidence level (e.g. 95% confidence level at $\alpha = 0.05$). For a more discussion on controlling *overall* probability of false rejection when testing at multiple voxels, please see §4.

2.1. *t*-test vs. *F*-test

Without distributional assumptions it is hard to compute tight bounds on the *p*-values[1], but under typical neuroimaging settings the normality assumption i.e. $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, is satisfied. Hence one can obtain the *p*-values by either assuming that the rows of $\tau\beta$ follow student-*t* distributions or that the residual, $\|Y - X\beta\|^2$, follows χ^2 distribution. The former assumption leads to *t*-tests and the latter to *F*-test. Below we discuss why we opt for *F*-tests over *t*-tests for SVRs.

For each independent row, i, of τ, the t statistic is computed as $t_0^i = \frac{\mathcal{T}^i\hat{\beta}}{\widehat{\mathrm{se}}(\mathcal{T}^i\hat{\beta})}$. For example, in Eq. (5), t_0^1 can be computed as:

$$t_0^1 = \frac{\hat{\beta}_1 - 2\hat{\beta}_2}{\widehat{\mathrm{se}}(\hat{\beta}_1 - 2\hat{\beta}_2)} = \frac{\hat{\beta}_1 - 2\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{11} + 4C_{22} - 4C_{12}}}, \quad (6)$$

[1] One might be able to compute some loose bounds using inequalities in convergence of random variables, such as Hoeffding's inequality [7].

[2] Camino is an open-source Diffusion-MRI processing.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $C_{ij} = cov(\beta_i, \beta_j)$. The key thing to note is that estimating C is not straightforward when we replace OLS with ϵ-SVR.

Since OLS minimizes least squared loss and is an unbiased estimator, the estimation error is equal to the variance in the estimation and the covariance matrix $C = (X^T X)^{-1}$. But not all loss functions and estimation procedures admit such clean bias-variance decomposition of estimation error [4] and computing C in such cases is not clear. The ϵ-insensitive L_1 loss cannot admit such a clear decomposition, although some interpretations could be made heuristically [11]. One can bootstrap and estimate variance but, not only is bootstrapping at each voxel in the brain computationally very expensive in a medical imaging setting, but also it is not clear how the bias can be estimated.

F-test allows us to infer statistical significances without needing the error decomposition and are based on residual sum of squares of the fit rather than the precision in the parameter estimates. Thus, one can perform inferences using a wide variety of estimators and loss functions. The two χ^2 variates that are needed to test the null hypotheses, H_0: $\tau\beta = 0$ are obtained as follows: We solve $\tau\beta = 0$ to represent the dependent coefficients using independent coefficients. Thus, we can obtain a *nested* reduced model $Y = Z\beta$. For example, for contrast in Eq. (5) we get:

$$\beta_1 = 2\beta_2 \text{ and } \beta_3 = \beta_4. \quad (7)$$

$$\therefore Y = \beta_0 + 2\beta_2 X_1 + \beta_2 X_2 + \beta_4 X_3 + \beta_4 X_4 = \beta_0 + \beta_2(2X_1 + X_2) + \beta_4(X_3 + X_4) = Z\gamma, \quad (8)$$

where $\beta = [\beta_0\ \beta_2\ \beta_4]$ and $Z = [\mathbf{1}\ (2X_1 + X_2)\ (X_3 + X_4)]$. Then we obtain the two residual sum of squares, $\|Y - X\hat{\beta}\|^2$ and $\|Y - Z\hat{\gamma}\|^2$ which form the two χ^2 variates used in computing the *F*-statistic. Note that for this particular contrast τ, we get a *t*-statistic for each independent row, but only one *F*-statistic. Thus, we lose specificity in terms of the effect of coefficients. However, in general, by carefully designing τ, one can achieve the desired specificity of effects even using *F*-tests.

2.2. Residual Degrees of Freedom

Both χ^2 and student-*t* distributions are parameterized by a degree-of-freedom. Since in our case the random variables are the residuals, they are called *residual* degrees of freedom (rdf).

In the case of a *t*-test using OLS, the rdf = n − rank(X) which equals $n - p$, when the explanatory variables are all linearly independent. H_0 is then rejected if $\forall i$, $|t_0^i| > t_{n-\text{rank}(X),1-\alpha}$, the critical value of *t* at $1 - \alpha$ significance level.

In the case of an *F*-test, the *F*-statistic is computed as:

$$F_0 = \frac{V_1/\text{rdf}_1}{V_2/\text{rdf}_2}, \text{ where} \quad (9)$$

$$V_1 = (\|Y - Z\hat{\gamma}\|^2 - \|Y - X\hat{\beta}\|^2) \sim \chi^2(\text{rdf}_1), \quad (10)$$

$$V_2 = (\|Y - X\hat{\beta}\|^2) \sim \chi^2(\text{rdf}_2). \quad (11)$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

H_0 is then rejected if $F_0 > F_{\text{rdf}_1,\text{rdf}_2,1-\alpha}$ the critical value of F at $1-\alpha$ significance level. Note that $F_0 > 0$, always. If OLS is used the degrees of freedom are computed as:

$$\begin{aligned}\text{rdf}_1 &= n-\text{rank}(Z)-(n-\text{rank}(X)) \\ &= \text{rank}(X)-\text{rank}(Z),\end{aligned} \quad (12)$$

$$\text{rdf}_2 = n-\text{rank}(X). \quad (13)$$

But when using ϵ-SVR the rdfs need to be computed differently. Let $\hat{Y} = HY$, where H is called the "hat matrix". In the case of *linear* ϵ-SVR, H can be obtained as:

$$\begin{aligned}\hat{Y} \equiv X\hat{\beta} = X\hat{\beta}\left[\tfrac{1}{n}(\mathbf{1}./Y)^T Y\right] = \tfrac{1}{n}\left[X\hat{\beta}(\mathbf{1}./Y)^T\right]Y, \\ \therefore H = \tfrac{1}{n}\left[X\hat{\beta}(\mathbf{1}./Y)^T\right].\end{aligned} \quad (14)$$

$\mathbf{1}./Y$ is just element-wise inversion of Y. Just for contrast, H in the case of OLS would be obtained as:

$$\begin{aligned}\hat{Y} \equiv X\hat{\beta} \text{ and } \hat{\beta} = (X^T X)^{-1} X^T Y, \\ \therefore \hat{Y} = X\left[(X^T X)^{-1} X^T\right] Y \;\&\; H = X(X^T X)^{-1} X^T.\end{aligned}$$

Notice the explicit dependence on Y, in case of ϵ-SVR. Once we have H, the rdf can be computed using Satterthwaite approximation [12, 6]:

$$\begin{aligned}\text{rdf} \equiv \text{tr}((I-H)^T(I-H)) = n-\text{tr}(2H-HH^T) \\ \approx n-1.25\text{tr}(H)+0.5.\end{aligned} \quad (15)$$

Thus using Eqs. (15) and (14), we can compute rdf_2. Now we need to compute rdf_1. Note that $V_1 = U_1 - U_2$, where

$$\begin{aligned}U_1 = \left|\left|Y - Z\hat{\gamma}\right|\right|^2 \sim \chi^2(\text{rdf}_{U_1}), \\ U_2 = \left|\left|Y - X\hat{\beta}\right|\right|^2 \sim \chi^2(\text{rdf}_{U_2}),\end{aligned}$$

and rdf_{U_1}, rdf_{U_2} can be computed similarly as rdf_2. Using Welch-Satterthwaite equation [17, 12], which can be used to approximate df of a linear combination of χ^2 variates, we can compute rdf_1 as:

$$\text{rdf}_1 \approx \frac{(U_1-U_2)^2}{U_1^2/\text{rdf}_{U_1} - U_2^2/\text{rdf}_{U_2}}. \quad (16)$$

3. Experiments

Diffusion tensor imaging (DTI) data sets from two different neuroimaging studies were used for our experiments to compare OLS and ϵ-SVR. DTI is a non-invasive method to characterize the microstructural properties and macroscopic organization of brain white

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

matter (WM) tissues [1]. The diffusion tensor is a positive-definite matrix that is a $2D$ manifold in $\mathbb{R}^3$. It captures the covariance of water diffusion in the three orthogonal Cartesian directions. Fractional anisotropy (FA), the most commonly used measure of diffusion anisotropy, is a normalized standard deviation of the eigenvalues that ranges between 0 and 1. The higher the value in a voxel, the more organized (in a primary direction) the WM is in that voxel. Below we present sample characteristics and data acquistion since the quantitative measures of DTI, such as FA, depend on the scanner and acquistion parameters.

(1) Autism Study—DTI data from 78 male subjects were used in this study: 42 high functioning subjects with autism spectrum disorders (ASD) and 36 Controls group-matched for age, handedness and IQ. DTI data were acquired on a Siemens Trio 3.0 Tesla Scanner with an 8-channel, receive-only head coil using a single-shot, spin-echo, EPI pulse sequence and SENSE parallel imaging (undersampling factor of 2). Diffusion-weighted images were acquired in 12 non-collinear diffusion encoding directions with diffusion weighting factor $b = 1000 s/mm^2$ in addition to a single reference image (b=0).

(2) Emotion-regulation Study—DTI data from 64 18-year-old adolescents were used. Cortisol (Cort) was obtained from salivary samples, collected over 3 consecutive days, when they were 4.5 years of age. Cort is an important steroid hormone implicated in the stress response, serving as an important measure in studies of emotion-regulation and anxiety [13]. The diffusion weighted images were acquired on a GE 3.0 Tesla scanner using 48 non-collinear diffusion encoding directions with diffusion weighting factor of $b = 1000 s/mm^2$ in addition to 8 $b = 0$ images. Eddy current related distortion and head motion of each data set were corrected using FSL software package ([14]) and distortions from field inhomogeneities were corrected using field maps.

For both the studies, the brain tissue was extracted using the brain extraction tool (BET), also part of the FSL [14]. The tensor elements were calculated using non-linear estimation using CAMINO[2]. It is important to establish spatial correspondence of voxels among all the subjects before performing VBA. State-of-the-art DTI registration toolkit DTI-TK[3] was used for spatially normalizing the subject data. All voxel based analyses were performed on spatially normalized $1 mm^3$ isotropic volumes with a final data resolution of $192 \times 224 \times 144$.

SurfStat-LIBSVM: We implemented the proposed max-margin general linear modeling by integrating two popular software packages: SurfStat [18] and LIBSVM [3]. This provides an effective MATLAB interface for neuroimaging studies. SurfStat allows intuitive programming of design and contrast matrices using higher level representations known as model formulas [18]. For example, if one wants to study the effect of group and age on fractional anisotropy (FA) in the brain by covarying for Gender, one could simply design the GLM as FA = 1 + Age + Group + Gender, where Age, Group and Gender are simply MAT-LAB arrays wrapped by a function called term.

Comparisons between OLS and ε-SVR: To demonstrate the advantage of using max-margin GLM, we examine the following two different GLMs, one for each study:

1. FA= β_0 + β_1Age + β_2Group + β_3Age * Group
2. FA= β_0 + β_1Cort + β_2Gender + β_3Cort * Gender

[3] http://www.nitrc.org/projects/dtitk

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where the corresponding null-hypotheses are $\beta_3 = 0$ for each of the model. A significant β_3 in the first model provides evidence that normal development of WM is different from development of WM in individuals with ASD. Similar interpretation can be made with alternate hypothesis from the second model. We estimate the above GLMs voxelwise at each voxel in the WM mask, which is defined as the set of voxels whose population-specific mean FA > 0.2. *F*-statistic maps computed using OLS and SVR for the two GLMs are shown in Figs. 1 and 2 respectively.

To visualize the scatter plots for the multilinear regressions ($p > 1$) we first need to "reduce" it to simple linear ($p = 1$) regression by removing the nuisance covariance (effect of variables not involved in the null-hypotheses) from the Y. For example, if $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3$ and we need a plot to show the effect of X_2 on Y, we first estimate β_0, β_1, β_2 and then estimate the simpler model, $\hat{Y}(=X\hat{\beta}) = \beta_0' + \beta_2' X_2$. We can then show the scatter plot of $\hat{Y}$ and X_2 along with line of slope β_2's and intercept β_0'. The scatter plots at one of the significant clusters ($F > 5$) in the cingulum are shown as insets in the Figs. 1 and 2. The $r^2 s \left(= 1 - \frac{\left| |Y - X\hat{\beta}| \right|^2}{\left| |Y - \bar{Y}| \right|} \right)$ in the scatter plots show the significance of the interactions in each group, while the *F*-stats on the *X*-labels show the significance of the group-difference in the respective interactions.

The distributions of the *F*-stats for the two models are shown in the log-scale in Fig. 3. The higher the *F*-stats, the better is the sensitivity of the estimation procedure to the underlying effect of a set of explanatory variables. ε-SVR has significant improvement both in terms of having more voxels with higher *F*-stats and fewer voxels with lower *F*-stats. One key difference in comparing OLS and SVR is that for latter, since H depends on Y, each voxel can have different rdf and hence we can have spatially varying degrees of freedom. In contrast, it is fixed per model when using OLS. Our computations indicate that the rdfs are almost constant (with small variance) across the white matter and hence we use the median of the rdfs in computing the *F*-stats. Finally, Fig. 4 shows the scatter plots of the quadratic effects of Age on the average FA in the cingulum cluster (from Fig. 1). Thus, GLM can also be used to investigate non-linear relationships by including non-linear terms in β. For all our experiments with SVR, we chose C to be the maximum of Y [9]. ε to be $0.6166 \times \sigma_Y$ [15].

4. Discussion and Future Directions

In this paper we presented a novel way of performing GLM based hypotheses testing by using ε-SVR, which we call max margin GLM. We demonstrated its potential advantages on real data from two different neuroimaging studies. We compared its performance with commonly used OLS. The improvements due to robustness of the estimation can be seen both in terms of obtaining higher *F*-statistics and also in addressing the nuisance covariance. While the robustness is mainly due to the ε-insensitive loss function, the improved *F*-stats are computable because of our approximation of the residual degrees of freedom using Welch-Satterthwaite approximation [17, 12]. The implementation is made by integrating popularly used software packages for a more direct impact of the presented work. To our best knowledge, this is the first attempt to apply a very successful loss function used in the machine learning community to the GLM framework for statistical significance testing of neuroimaging data.

We foresee three main lines of future work for the proposed work. (1) In our experiments, although SVR produced higher *F*-stats compared to OLS, the stats were not above the threshold of Bonferoni (BON) correction ($F_{\mathrm{rdf}_1, \mathrm{rdf}_2, 1-\frac{\alpha}{v}}$) for the multiple comparisons

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

problem [10]. The multiple-comparisons problem has similarities to generalization and over-fitting problems faced in machine learning. There exist several approaches beyond näive Bonferonni correction such as random field theory (RFT) based correction, false discovery rate (FDR) control [2] and permutations-based correction. Based on upper bounds on some topological properties obtained using algebraic geometry theory, and treating $\{F_0(x)\}_{x=1}^{v}$ as a statistical field, RFT proposes to correct using the following approximation:

$$\alpha_{\mathrm{RFT}} = p(\max_x F_0(x) > h) \approx R \times EC(\mathcal{F}_h),$$

where R is the "resel" (resolution element) count, EC denotes the Euler characteristic of a set and $\mathcal{F}_h = \{x: F_0(x) > h\}$ is called the "excursion set". Usually, if h is reasonably large (say $> KF_{\mathrm{rdf}_1,\mathrm{rdf}_{2,1-\alpha}}$), then $\alpha_{\mathrm{RFT}} \approx \alpha_{\mathrm{BON}} (= v\alpha)$ and hence the F tests can be more sensitive even with a corrected threshold. By actually using spatially varying rdf (instead of just median), whether SVR can potentially result in more favorable $\mathcal{F}_h$ comparable to OLS, is an interesting question. (2) Although we compared our method with OLS, the most widely used estimation in neuroimaging setting, comparison with other robust estimation techniques such as iterative weighted least squares and other ridge regression methods would throw more light on the advantages of the SVR in this setting. (3) Extending this work to a multivariate inference setting ($Y \in \mathbb{R}^{n\times K}$) by combining better loss functions from multi-task learning and multivariate hypotheses testing is also a very interesting direction of future work. Finally, relaxing the normality assumptions and estimating distribution-free p-values using results from convergence of random variables (e.g., [7]) would help significance testing of not just SVR based estimation but also OLS in the context of GLM.

References

1. Basser P, Mattiello J, Bihan D. Estimation of the effective self-diffusion tensor from NMR spin echo. J Magn Reson. 1994; 103:247–254.
2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc. 1995; 57(1):125–133.
3. Chang, C.; Lin, C. LIBSVM: a lib for SVMs. 2001.
4. Domingos P. A unified bias-variance decomposition and its applications. Proc ICML. 2000:231–238.
5. Gunter L, Zhu J. Efficient computation and model selection for the support vector regression. Neural Computation. 2007; 19:1633–1655. [PubMed: 17444762]
6. Hastie, T.; Tibshirani, R. Generalized additive models. CRC Press; 1990.
7. Hoeffding W. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association. 1963; 58(301):13–30.
8. Huber, PJ. Robust Statistics. John Wiley & Sons; NY: 1981.
9. Mattera and Haykin. Advances in Kernel Methods. The MIT Press; 1999.
10. Miller, R. Simultaneous Statistical Inference. Springer Verlag; New York: 1981.
11. Park J, Kim J. Quant. reg. with an ε-insens. loss in a rep. ker. Hilbert space. Stat & Prob Let. 2011; 81(1):62–70.
12. Satterthwaite F. An approximate distribution of estimates of variance components. Biometrics Bul. 1946; 2:110–114.
13. Smider N, Essex M, Kalin N, et al. Salivary cortisol as a predictor of socioemotional adjustment during kindergarten: a prospective study. Child Dev. 2002; 73:75–92. [PubMed: 14717245]
14. Smith S, et al. Advances in func. & struc. MR img. analysis & implementation as FSL. NIMG. 2004; 23:208–219.
15. Smola A, Murata N, et al. Asymptotically opt. choice of ε-loss for SVMs. ICANN. :105–110.
16. Smola, A.; Schölkopf, B. A Tutorial on SVR. 2003.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

17. Welch B. The generalization of student's prob. when several diff. population var. are involved. Biometrika. 1947:28–35. [PubMed: 20287819]

18. Worsley K, et al. Surfstat: A MATLAB toolbox for stat. anal. of univar. & multivar. surf. & vol. data using lin. mixed effects models & RFT. NIMG. 2009; 47:102–102.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

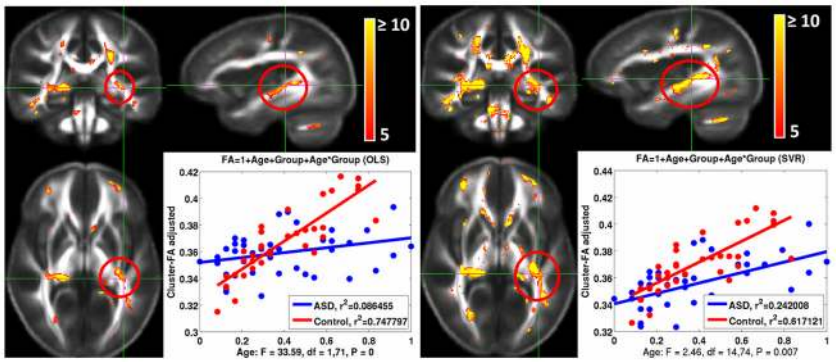

Figure 1.

F-statistic maps using OLS (left) and SVR (right) overlaid on the corresponding mean FA maps. The GLM (shown on top of the scatter plots) aims at measuring the group difference between ASD and Controls in terms of interaction between Age and FA. It can be observed that both OLS and SVR show similar regions of significance but the SVR has higher *F*-stats. The scatter plots show the regression between the avg. FA (adjusted for the Age and Group as nuisance covariates) in the encircled cluster on the cingulum bundle and Age. SVR not only obtains higher *F*-stats but also seems to account for the nuisance covariance (especially in the ASD group) more accurately as can be seen in the scatter plots.

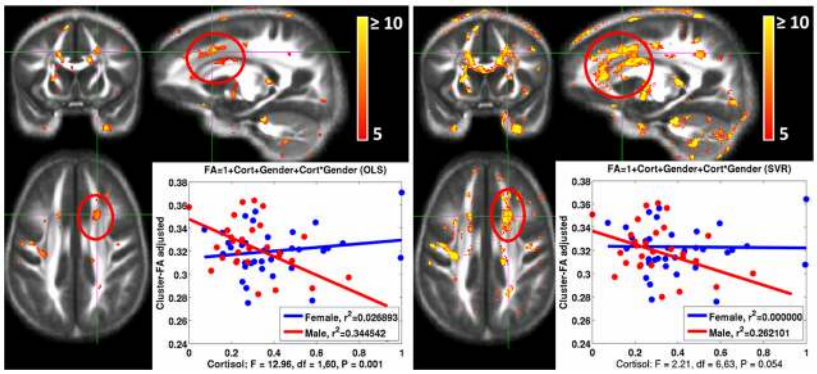

Figure 2.

Similar to Fig. 1 but the GLM is to measure effect of Cort and Gender interactions on FA. We can observe that SVR produces higher *F*-stat maps and in spatially more contiguous regions (encircled in red) thus enabling biologically more meaningful results. The scatter plots for both OLS and SVR are also shown for a cluster in the superior-frontal projections of white matter tracts.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

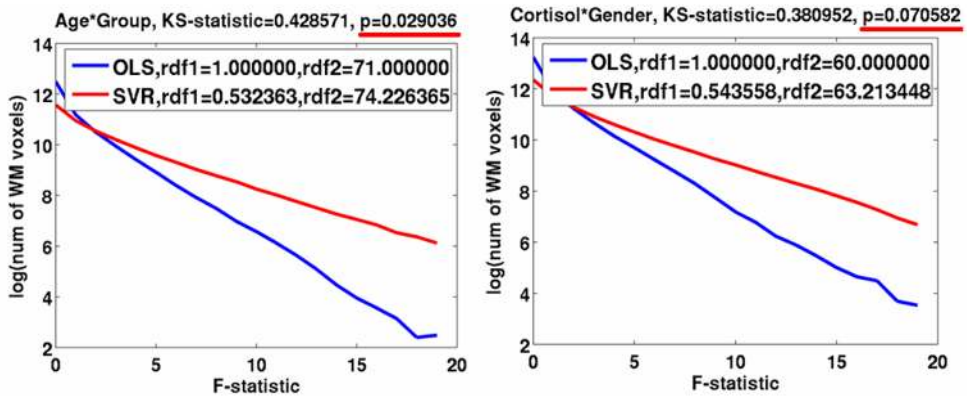

Figure 3.

Log distributions of the *F* stats of the WM voxels for the two GLMs shown in Figs. 1,2. We can observe that SVR produces improved *F*-stats in a statistically significant way (using Kolmogorov-Smirnov test). The numerator (rdf1) and denominator (rdf2) degrees of freedom for both OLS and SVR are shown in the legends.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

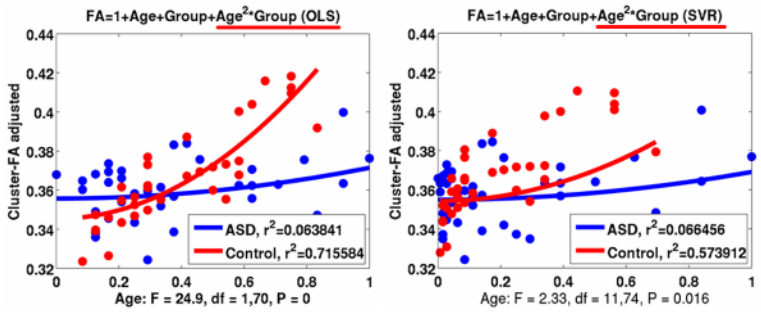

Figure 4.

The quadratic effect of Age on the adjusted avg. FA of the cingulum cluster from Fig. 1. We can observe that after adjusting for the variance of the linear and quadratic terms, the group-difference in the quadratic effects although significant ($p = 0.016$), is reduced when using SVR compared to OLS. This shows that SVR can better account for variance of nuisance variables.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript