Improving predictive models using non-spherical Gaussian priors

Anwar O. Nunez-Elizalde* **(aone@berkeley.edu)**

Alexander G. Huth* **(alex.huth@berkeley.edu)**

Jack L. Gallant (gallant@berkeley.edu)
Helen Wills Neuroscience Institute, UC Berkeley
Berkeley, CA 94720 USA

Abstract

Predictive models for neural or fMRI data are often fit using regression methods that employ priors on the model parameters. One widely used method is ridge regression, which employs a Gaussian prior that has equal and independent variance for all parameters (i.e. a spherical multivariate Gaussian). However, a spherical prior is not always appropriate: there are many cases where expert knowledge or hypotheses about the structure of the model parameters could be used to construct a better prior. Here we show that Tikhonov regression with non-spherical Gaussian priors can improve several predictive models for fMRI data. This is particularly important when combining feature spaces into a single predictive model. Finally, we demonstrate a computationally efficient method for incorporating non-spherical priors into regression models.

Keywords: encoding models; predictive models

Introduction

Non-spherical Gaussian priors can be employed using a generalized form of ridge known as Tikhonov regression. Yet Tikhonov regression is not (explicitly) used very often in neuroscience. Here we show that Tikhonov regression with non-spherical Gaussian priors can improve several predictive models for fMRI data. We also show that many earlier studies have implicitly used Tikhonov regression by linearly transforming the regressors before performing ridge regression. Our first result is based on an fMRI experiment in which subjects listened to several hours of natural stories (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). A predictive model in which the features were indicator variables for each unique word provided poor predictions of held-out data. However, performance was improved drastically by employing a non-spherical Gaussian prior in which words that have similar meanings are likely to be assigned similar weights. Our second result comes form an fMRI experiment in which subjects watched hours of natural movies (Nishimoto et al., 2011). A predictive model in which low-level structural and high-level categorical features were combined predicted held-out responses to novel stimuli poorly when using a spherical prior. We show that a non-spherical prior where the two feature spaces have unequal variance performs better.

Methods

The key insight underlying our results is that Tikhonov regression problems can be converted to a standard form (Hansen, 1998) that is more interpretable and yields more computationally efficient solutions than the traditional formulation. The traditional solution for Tikhonov regression is given by: $\beta = \left(X^\top X + \lambda C^\top C\right)^{-1} X^\top Y$, where C is the penalty matrix, and the prior distribution on β is multivariate normal with zero mean and covariance equal to $\left(\lambda C^\top C\right)^{-1}$. This traditional form is expensive to compute. However, by a change of variables, $A = XC^{-1}$, the problem is converted to standard form: $\beta = C^{-1}\left(A^\top A + \lambda I\right)^{-1} A^\top Y$. Thus, we see that Tikhonov regression in standard form is identical to ridge regression after projecting the regressors onto the inverse of the penalty matrix. The matrices are simultaneously diagonalizable, so we can efficiently test different values of without fully recomputing any matrix inverse. Furthermore, this shows that any model that uses ridge regression after applying a linear transformation to the regressors is implicitly doing Tikhonov regression. This suggests alternative interpretations of the models used in some previous studies For example, the model used in Huth, et al., 2012, supplemented manually labeled visual categories (e.g. dog) with higher-level categories inferred using WordNet (e.g. canine, carnivore, etc.). Because this operation is linear, the model can be re-interpreted as Tikhonov regression where the prior covariance between two parameters (e.g. dog and cat) is proportional to the number of higher-level categories that they have in common.

Results

Language Experiment. We recorded BOLD fMRI data from two subjects while they listened to hours of natural narrative stories (Huth et al., 2016). The stories were annotated with the exact time and identity of each spoken word. We first fit a simple model that predicts BOLD timecourses in each voxel as a weighted sum across word timecourses, using a spherical prior on the weights. To extend this model we incorporated a non-spherical semantic prior, assuming that words having similar meanings should have similar weights. Semantic similarity was determined using a word embedding space based on word co-occurrence statistics across a large corpus of text (Wehbe et al., 2014). Models based on the spherical and semantic priors were fit to one set of data and then tested on a separate, held-out dataset. We found that the non-spherical semantic prior outperformed the spherical prior by a large margin. This shows that theoretically motivated non-spherical priors can improve model performance.

Movie Experiment. We recorded BOLD fMRI data from three subjects while they watched natural movies. Motion-energy features were extracted from these movies using a spatio-temporal Gabor pyramid (Nishimoto et al., 2011), and visual object and action categories present in each one second segment of the movies were labeled by hand (Huth, et al., 2012). We fit a single model that predicts BOLD responses as a linear combination of both motion-energy and category features. Applying a spherical prior to this combined model yielded poor predictions. To address this issue we fit the same combined model using a non-spherical prior in which all parameters are independent but the variance is different for the two feature spaces. This imposes one spherical prior on the motion-energy features and a different spherical prior on the category features. Models based on the spherical and non-spherical priors were tested on a held-out dataset. The non-spherical prior provided far better predictions than the spherical prior (Fig. 1). This suggests that non-spherical priors are vital for optimally combining different feature spaces into a single predictive model.

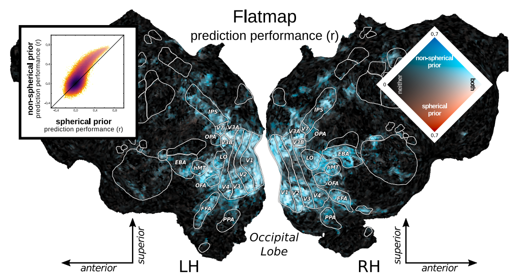

Figure 1: Movie experiment. Comparison of prediction accuracy for the same model fit with different priors. *Left inset*: Points that lie above the diagonal line correspond to voxels that were more accurately predicted with a non-spherical prior. *Flatmap*: The prediction accuracy of the same model fit with a spherical (red) and a non-spherical (blue) prior is shown on a flattened view of the cortical sheet. The model fit with a non-spherical prior provides better predictions in most of visual cortex.

References

Hansen, P. C. (1998). *Rank-deficient and discrete ill-posed problems* (Vol. 4). Siam.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PloS one*.