Stimulus Identification from fMRI scans

Charles Zheng and Yuval Benjamini

Stanford University, Department of Statistics

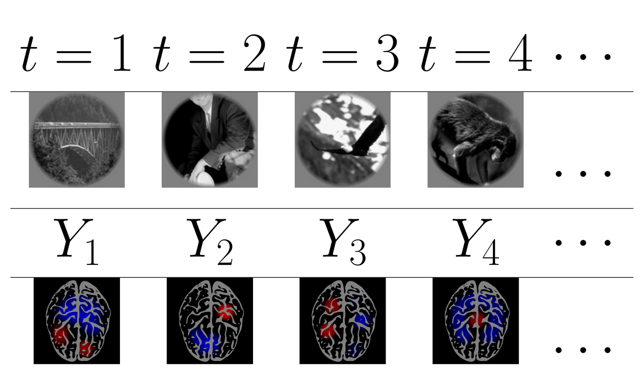
Overview

Seeking to explain the processes behind human perception, scientists employ *forward models* to model the causal relationship between stimulus and neural activity. But how can we measure the quality of these models? Kay et al (2008) introduced the task of *identification* as a way to demonstrate the fidelity and generalizability of the model.

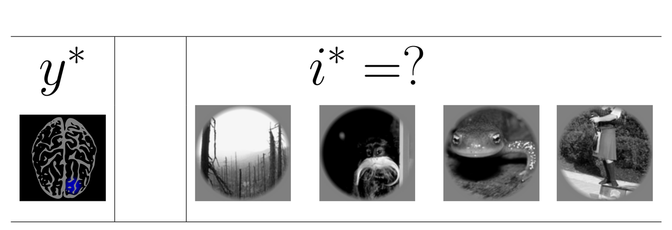
Using the data of Kay *et al.* as a motivating example, we consider the statistical problem of optimal identification. While identification superficially resembles a classification task (with many classes), it combines the challenge of multivariate regression with high-dimensional discrimination.

Data

- Sequence of stimuli (pictures) shown at time $t = 1, \ldots, T = 3500$
- Record subject's multivariate response $y_t \in \mathbb{R}^p$, here $p \approx 20000$



Identification



- Let S be a set of stimuli, possibly outside the training set! $|S|$ can range from 120 to 10000
- Scientist picks a stimulus i^* from S and measures the subject's reponse y^*
- Can the statistician *identify* $i^* \in S$ from y^*?

Remark. In order to identify images outside the training set, we need some way to generalize beyond the training set!

Previous work

Previous work [1][2] generally follows likelihood-based approaches. Consider a parametric model

$$Y \sim F_\theta(X)$$

where $Y_{T\times p}$ is a matrix containing the T of recorded responses, and where $X_{T\times q}$ is the matrix of the *image features* of the corresponding stimuli. E.g. columns of X are Gabor filters with $q \approx 10000$, and θ is some parameter to be estimated. Let $x_i : i \in S$ denote features of the test stimuli, and identify y^* based on the maximum likelihood (ML) principle

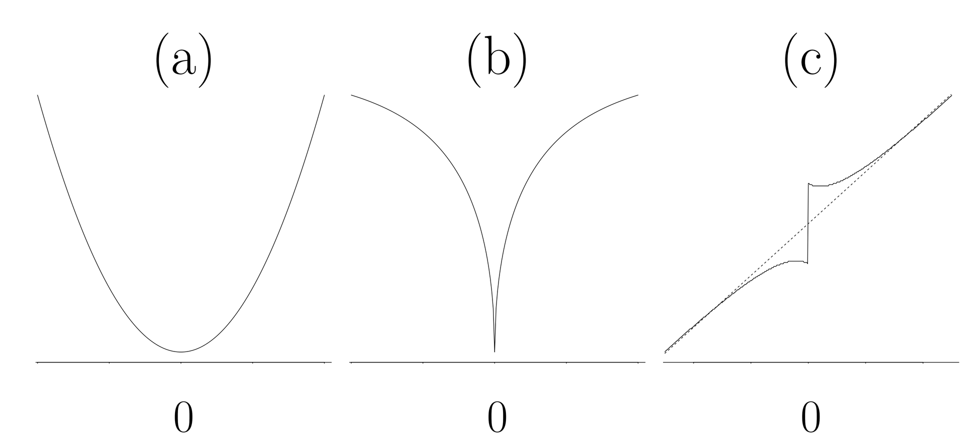
$$i^* = \text{argmax}_i \ell(y^*, x_i)$$

We take the following as a representative approach

- Assume the normal mutivariate linear model
$$Y \sim N(XB, \Sigma_E^2)$$
- Obtain point estimates of B and Σ_E. E.g. B estimated using elastic net [4], and
$$\hat{\Sigma}_E = (1-\alpha)\hat{\text{Cov}}(Y - \hat{Y}) + \alpha \text{diag}(\hat{\text{Cov}}(Y - \hat{Y}))$$
where $\hat{\text{Cov}}$ = sample covariance, $\alpha \in (0, 1)$.
- Identify the stimulus i^* by
$$i^* = \text{argmin}_i (x_i^T B - y^*)^T \hat{\Sigma}_E^{-1} (x_i^T B - y^*)$$

Limitations of ML

- Point estimates obtained by minimizing *prediction error*, but the loss function for identification is different!
- In fact, any estimate of B which is degenerate (identical columns up to scaling) is suboptimal



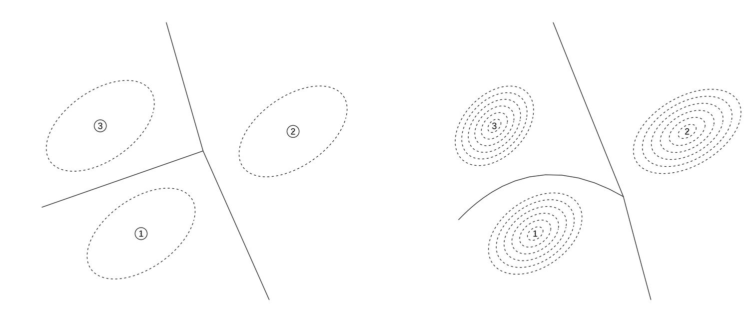
(a) Squared error loss (vertical axis) and *(b)* loss function for identification, as a function of the difference between the true mean signal and the predicted signal.

(c) The optimal point estimate for identification (solid) vs the optimal point estimate for regression (dashed) diverge sharply at $B = 0$ in the one-dimensional case. The same phenomenon occurs for higher dimensions when B is degenerate.

Empirical Bayes

Can we find a principled alternative to ML?

- *Idea*: Unlike ML, the Bayes rule surely optimizes the "correct" objective function
- *Problem*: We don't know the hyperparameters for Bayesian inference
- *Emprical Bayes*: use the data to estimate the covariances Σ_B and Σ_E, then compute posterior distribution of B
- In contrast to ML, which results in *linear decision boundaries* (below: left), Empirical Bayes (EB) results in *quadratic boundaries* (below: right)



ML EB

Technical details

Model

- Noise $E_t \sim N(0, \Sigma_E)$ iid
- Coefficients $B_i \sim N(0, \sigma_i^2 I)$ for $i = 1, \ldots, p$

Estimate hyperparameters

- Use *eigenprism* (Janson 2015) to estimate $\theta_i^2 = ||B_i||^2$ for $i = 1, \ldots, p$
- Set $\sigma_i^2 = \hat{\theta}_i^2 / q$
- Estimate $\hat{B}$ as posterior mean
- Estimate Σ_E (same as in ML)

Compute posterior

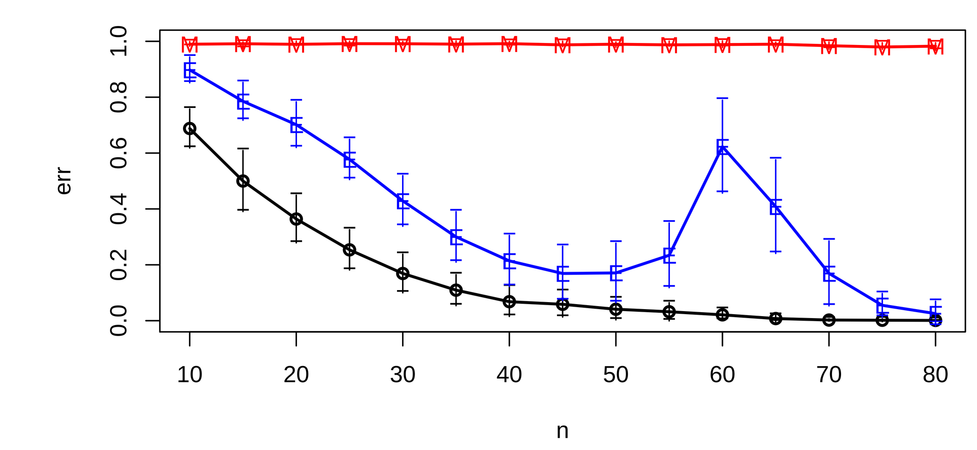
- Closed-form expressions for posterior of B
- Computational bottleneck: inverting a $pq \times pq$ covariance matrix

Apply Bayes rule

- *Uncertainty* in B is reflected as *added noise*
- Result: posterior $\text{Cov}(y^*|i^*)$ varies, hence *quadratic boundaries*

Simulation Results

- Parameters $p = q = 60$, random B and Σ_E
- Empirical bayes outperforms ML when $n < q$... however, still unstable!



(E) Empirical Bayes, (M) Maximum likelihood, (o) Bayes risk *(knowing true Σ_B, Σ_E)*

Ongoing Work

- Why does error *increase* with sample size!? Refine covariance estimation methods..
- Required cost of $O((pq)^3)$ unacceptable for real data... develop tractable approximations

Conclusions

- ML-based approaches rely on point estimates, and hence optimize the wrong objective function
- Empirical bayes achieves better performance by approximating the Bayes rule, but the "empirical" part remains tricky
- Better theoretical understanding is needed to explain why EB succeeds (and sometimes fails)

References

[1] Kay et al. *Nature* (2008)
[2] Vu et al. *Annals of Applied Statistics* (2011)
[3] Janson et al. (2015) http://arxiv.org/abs/1505.02097
[4] Zou et al. *J. R. Statist. Soc. B* (2005)

Acknowledgements

This work was supported by an NSF graduate research fellowship. We are also grateful to the travel support provided by the SAND 7 conference.