

MATH1298 Chapter 5: Model Selection and Evaluation

Additional material: Feature selection under presence of interaction terms - a very brief coverage

Vural Aksakalli & Yong Kai Wong

4 Oct 2018

Background

- A motivating example: do gene interactions help predict cancer type?
- How to determine interactions in high dimensions *optimally*?
- Focus on pairwise interaction terms (no quadratic terms)
- Given p features with 2-way pairwise interaction terms, the number of possible features is

$$p + \binom{p}{2} = \frac{p(p+1)}{2}$$

- Side question: what is the number of possible features if we include quadratic terms?
- **Not in examination**

Notations and terminology

- Y : response feature
- X_j : explanatory feature j for $j = 1, 2, \dots, p$
- Model formulation:

$$g(Y) = \beta_0 + \sum_{j=1}^p \beta_j X_j + \sum_{i < j} \beta_{i:j} X_j X_i$$

- A precise definition [Lim and Hastie, 2018, p.1]:

“When a function $f(x_1, x_2)$ cannot be expressed as $h_1(x_1) + h_2(x_2)$ for some functions h_1 and h_2 , we say that there is an interaction in f between x_1 and x_2 .”

- Introduction of “hierarchy”

Terminology: hierarchy

Lim and Hastie [2018] define:

Hierarchy	Description
Strong	Interactions are only among pairs of nonzero main effects
Weak	Each interaction has only one of its main effects present
Anti-hierarchical	Interactions are only among pairs of main effects that are not present
Pure interaction	No main effects present; only interactions

Terminology: an example of hierarchy

Consider three explanatory features: $\mathbf{X} = \{X_1, X_2, X_3\}$

- ① Strong hierarchy: $\{X_1, X_2, X_1X_2\}$
- ② Weak hierarchy: $\{X_1, X_1X_2, X_1X_3\}$
- ③ Anti-hierarchical: $\{X_2, X_1X_3\}$
- ④ Pure interaction: $\{X_2X_3, X_1X_3, X_1X_2\}$

In practice, how do we know? How can we detect?

Main methods to identify pairwise interaction terms

- 1 Statistical hypothesis test [Cox, 1984]
- 2 Regularisation, e.g. LASSO
- 3 Wrapper Feature Selection [Kohavi and John, 1997]

Statistical hypothesis test: an example

Considers two models:

① $g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

② $g(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{1:2} X_1 X_2$

Run hypothesis on test on $\beta_{1:2} = 0$.

Regularisation: glinternet

- Let $l(y_i; \beta)$ be the **negative** log-likelihood contribution
- Elastic-net [Friedman et al., 2010]

$$\arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n l(y_i; \beta) + \lambda[(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1]$$

- glinternet [Lim and Hastie, 2018]: group-based LASSO ($\alpha = 1$) by imposing additional constraints on $\beta_{k:j}$ and β_j
- glinternet package is available in R

Wrapper feature selection

Extending feature selection by including interaction terms

- ① SFFS: Sequential (Floating) Forward Selection [Pudil et al., 1994]
- ② ~~SFBS: Sequential (Floating) Backward Selection~~ [1994]
- ③ SP-FSR [2016]

SP-FSR algorithm

- Introduced by Aksakalli and Malekipirbazari [2016]
- Based on Simultaneous Perturbation Stochastic Approximation [Spall, 1992]
- Refined by Yenice et al. [2018]
- Pseudo gradient descent method on the loss function
- spFSR package is available in R [Aksakalli et al., 2018]

SP-FSR algorithm to identify interactions

- Assume a strong hierarchy
- Simplified version of **two-step SP-FSR** algorithm:
 - 1 Identify the optimal set of k main effects using SP-FSR
 - 2 Search k' , number of interactions from k main effects with SP-FSR
- k and k' can be determined via grid search or automatically
- GitHub link: <https://github.com/yongkai17/n-way-spFSR>

Limitations

Both algorithms

- Only two-way interaction terms under strong hierarchy
- Both incorporate cross-validation. Why is this a limitation?

`glinternet`

- Applicable to regression and logistic classification problems only.
- Not available for log-log and probit regression models.

SP-FSR algorithm

- Not available for Poisson regression which is not available in `mlr`
- Implementation in R hence slow
- Not experiment with other binary models yet.

References I

- Vural Aksakalli and Milad Malekipirbazari. Feature selection via binary simultaneous perturbation stochastic approximation. *Pattern Recognition Letters*, 75(Supplement C):41 – 47, 2016. ISSN 0167-8655. doi: <https://doi.org/10.1016/j.patrec.2016.03.002>. URL <http://www.sciencedirect.com/science/article/pii/S0167865516000702>.
- Vural Aksakalli, Babak Abbasi, and Yong Kai Wong. *spFSR: Feature Selection and Ranking by Simultaneous Perturbation Stochastic Approximation*, 2018. <https://www.featureranking.com/>, <https://arxiv.org/abs/1804.05589>.
- D. R. Cox. Interaction. *International Statistical Review / Revue Internationale de Statistique*, 52(1):1–24, 1984. URL <http://www.jstor.org/stable/1403235>.

References II

- Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, 33(1):1–22, 2010. ISSN 1548-7660. doi: 10.18637/jss.v033.i01. URL <https://www.jstatsoft.org/v033/i01>.
- R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 1(2):273–324, 1997.
- Michael Lim and Trevor Hastie. *glinternet: Learning Interactions via Hierarchical Group-Lasso Regularization*, 2018. URL <https://CRAN.R-project.org/package=glinternet>. R package version 1.0.7.
- P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler. Floating search methods for feature selection with nonmonotonic criterion functions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*, volume 2, pages 279–283 vol.2, Oct 1994. doi: 10.1109/ICPR.1994.576920.

References III

- James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE*, 37(3):322–341, 3 1992.
- Zeren D. Yenice, Niranjana Adhikari, Yong Kai Wong, Vural Aksakalli, Alev Taskin Gumus, and Babak Abbasi. Spsa-fsr: Simultaneous perturbation stochastic approximation in feature selection and ranking. 2018. URL <https://arxiv.org/abs/1804.05589>.