Abstracts of the Ninth Workshop on Simulation (2018) 1-2

Algebraic views on classification problems

Gherardo Varando [1] and Eva Riccomagno [2]

1 Introduction

We study generative classifiers for binary class over categorical predictors, that is models of the joint probability distribution $\mathrm{P} > 0$ over the predictors $\mathbf{X} \in \boldsymbol{\mathcal{X}}$ and the class variable $C \in \{-1, +1\}$. Every generative classifier induce a discrimination function,

$$f_{\mathrm{P}} = \ln(\mathrm{P}(\mathbf{X}, C = +1)) - \ln(\mathrm{P}(\mathbf{X}, C = -1)),$$

such that the maximum a posteriori prediction $\arg\max_{c \in \{-1,+1\}} \mathrm{P}(C = c|\mathbf{X})$ is equal to the sign of f_{P}.

It is known that the form of the induced function f_{P} is connected to the conditional independece assumptions that hold in P [5, 4, 6]. For example the naive Bayes assumtpion $(X_i \perp\!\!\!\perp X_j | C)$ translates, for the discrimination functions, in the following decomposition,

$$f_{\mathrm{P}}(x_1, \dots, x_n) = \sum_i f_i(x_i). \tag{1}$$

Complementarily we present a study of the set of generative classifier such that their induced functions satisfie the factorization in Equation (1).

$$\mathcal{P}_{\emptyset} = \{\mathrm{P} > 0 \text{ s.t. } f_{\mathrm{P}} = \sum_i f_i(x_i)\}.$$

2 Constant interactions models

Consider generative classifiers over two binary predictor variables X_1, X_2 and define the odds ratio of the conditional distribution of the predictors given the class variable,

$$\alpha[\mathrm{P}(X_1, X_2 | C = c)] = \frac{\mathrm{P}(X_1 = 0, X_2 = 0 | C = c)\, \mathrm{P}(X_1 = 1, X_2 = 1 | C = c)}{\mathrm{P}(X_1 = 1, X_2 = 0 | C = c)\, \mathrm{P}(X_1 = 0, X_2 = 1 | C = c)}.$$

We can prove the following equivalence that characterize the set $\mathcal{P}_{\emptyset}$.

$$\mathrm{P} \in \mathcal{P}_{\emptyset} \Leftrightarrow \alpha[\mathrm{P}(X_1, X_2 | C = +1)] = \alpha[\mathrm{P}(X_1, X_2 | C = -1)].$$

[1] E-mail: gherardo.varando@gmail.com

[2] Dipartimento di Matematica, Universitá degli Studi di Genova, Genova, Italy, E-mail: riccomagno@dima.unige.it

As Fienberg [2] we consider the manifold of costant interaction as the probabilities with odds ratios equal to $\alpha > 0$.

$$\mathcal{M}(\alpha) = \{\mathrm{Q} > 0 \text{ s.t. } \alpha[\mathrm{Q}] = \alpha\}.$$

If we parametrize a generative classifier $\mathrm{P} = \mathrm{P}(C)\,\mathrm{P}(X_1, X_2|C)$ we have that,

$$\mathrm{P} \in \mathcal{P}_\emptyset \Leftrightarrow \mathrm{P}(X_1, X_2|C = \pm 1) \in \mathcal{M}(\alpha),$$

for some $\alpha > 0$.

Obviously, naive Bayes classifiers belong to $\mathcal{P}_\emptyset$, in particular they correspond to the choice $\alpha = 1$ that reduces $\mathcal{M}(1)$ to the manifold of independence [3, 1].

The above characterization can be extended to more than two categorical predictors, and generalizing the odds ratios we can similarly consider more complex factorizations of the descrimination function f_P.

Moreover models in $\mathcal{P}_\emptyset$ can be seen as generative classifiers equivalent to the logistic regression and thus we investigate maximum-likelihood estimation over $\mathcal{P}_\emptyset$.

References

[1] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics*. Oberwolfach Seminars. Birkhäuser Basel, 2009.

[2] Stephen E. Fienberg. The geometry of an $r \times c$ contingency table. *Ann. Math. Statist.*, 39(4):1186–1190, 08 1968.

[3] Stephen E. Fienberg and John P. Gilbert. The geometry of a two by two contingency table. *Journal of the American Statistical Association*, 65(330):694–701, 1970.

[4] Manfred Jaeger. Probabilistic classifiers and the concepts they recognize. In Tom Fawcett and Nina Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning (ICML-03)*, pages 266–273. AAAI Press, 2003.

[5] Marvin Minsky. Steps toward artificial intelligence. In *Computers and Thought*, pages 406–450. McGraw-Hill, 1961.

[6] Gherardo Varando, Concha Bielza, and Pedro Larranaga. Decision boundary for discrete bayesian network classifiers. *Journal of Machine Learning Research*, 16:2725–2749, 2015.