



VI Congreso de Jóvenes Investigadores

Real Sociedad Matemática Española

León, Febrero de 2023

Fundamentos Matemáticos del Aprendizaje Semi-Supervisado

A. Daniel Sanz-Alonso*

En esta charla plantearé algunos problemas matemáticos y estadísticos que subyacen la formulación Bayesiana del aprendizaje semi-supervisado [2]. El objetivo de esta modalidad de aprendizaje es asignar etiquetas a un gran número de datos usando pocos ejemplos. En la formulación Bayesiana, la medida posterior sobre la función de etiquetado combina la verosimilitud de los ejemplos con una medida *a priori* que actúa como regularización. Describiré cómo definir la medida *a priori* utilizando procesos Gaussianos en grafos [5], las propiedades de concentración de la medida posterior [6], y técnicas de muestreo cuya tasa de convergencia no deteriora con el número de datos a etiquetar [1]. Finalmente, mencionaré ejemplos de otras aplicaciones donde ideas parecidas pueden ser útiles, como en la solución de problemas inversos [3, 4] y optimización [7] en variedades.

Agradecimientos La investigación resumida en esta presentación ha contado con el apoyo de las siguientes instituciones: Department of Energy, National Geospatial Intelligence Agency, National Science Foundation, Fundación BBVA.

Referencias

- [1] N. Garcia Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso. *On the consistency of graph based Bayesian learning and the scalability of sampling algorithms*. Journal of Machine Learning Research. **21** (2020) 1-47.
- [2] N. Garcia Trillos, D. Sanz-Alonso, and R. Yang. *Mathematical foundations of graph-based Bayesian semi-supervised learning*. Notices of the American Mathematical Society. **10** (2022) 1717-1729.
- [3] J. Harlim, S. Jiang, H. Kim, and D. Sanz-Alonso, R. Yang. *Graph-based prior and forward models for inverse problems on manifolds with boundaries*. Inverse Problems. **38** (2022) 1-28.
- [4] J. Harlim, D. Sanz-Alonso, and R. Yang. *Kernel methods for Bayesian elliptic inverse problems on manifolds*. SIAM/ASA Journal on Uncertainty Quantification. **8** (2022) 1414-1445.
- [5] D. Sanz-Alonso and R. Yang. *The SPDE approach to Matérn fields: graph representations*. Statistical Science. **37** (2022) 519-540.
- [6] D. Sanz-Alonso, and R. Yang. *Unlabeled data help in graph-based semi-supervised learning: a Bayesian nonparametrics perspective*. Journal of Machine Learning Research. **23** (2022) 1-28.
- [7] H. Kim, D. Sanz-Alonso, and R. Yang. *Optimization on manifolds via graph Gaussian processes*. Preprint (2022).

*Department of Statistics, University of Chicago, Chicago, IL 60637, Valencia Email: sanzalonso@uchicago.edu