Kernel methods for Earth Observation: physics, inverse modeling and causality

Daniel Hernández-Lobato
Machine Learning Group (GAA)
Universidad Autónoma de Madrid, Madrid, Spain

✉ daniel.hernandez@uam.es | 🌐 http://dhnzl.org

KERMES: Advances in kernel methods for structured data, TEC2016-81900-REDT
KOM, 11/07/17
Faculty Members:
- José Ramón Dorronsoro Ibero (Full Professor)
- Alberto Suárez González (Full Professor)
- Gonzalo Martínez Muñoz (Associate Professor)
- Daniel Hernández-Lobato (Lecturer)
- Ana González Marcos (Lecturer)

PhD. Students:
- Eduardo César Garrido Merchán
- Carlos Villacampa Calvo
- María Cortés Alonso
- Alberto Torres Barragán
- Maryam Sabzevari
- David Díaz Vico

Post-docs:
- Carlos Alaíz Gudín
- Ángela Fernández Pascual
- Irene Rodríguez Luján

Master students:
- Gonzalo Hernández Muñoz
- Carlos Ramos Carreño
- Alejandro Catalina Feliú

http://www.eps.uam.es/%7egaa
Research Lines: Large Scale Gaussian Process Classification

- Binary and Multi-class **classification** with GPs.
- Very large datasets with **millions** of instances.
- **Expectation propagation** as an alternative to VI methods.
- Training via **mini-batches** and **stochastic gradients**.

![Graph showing Airline Delays](image-url)

**Methods**
- EP
- Linear
- SEP
- VI

**Axes**
- **X-axis**: Training Time in Seconds in a Log10 Scale
- **Y-axis**: Test Error
- Binary and Multi-class classification with GPs.
- Very large datasets with millions of instances.
- Expectation propagation as an alternative to VI methods.
- Training via mini-batches and stochastic gradients.
More **flexible** and give better **uncertainty estimates**.
- Automatically **learn a kernel** that works well for the data.
- **Alleviate the problems** of sparse GP approximations.
- Can be trained using **stochastic gradients** and **mini-batches**.

\[
\begin{align*}
\text{More flexible and give better uncertainty estimates.} \\
\text{Automatically learn a kernel that works well for the data.} \\
\text{Alleviate the problems of sparse GP approximations.} \\
\text{Can be trained using stochastic gradients and mini-batches.}
\end{align*}
\]
Efficient optimization of expensive **black-box** functions.
- Evaluations and gradients contaminated with noise.
- Very useful tool for **finding good model hyper-parameters**.
- Allows for **inequality constraints** and several objectives.
Efficient optimization of expensive black-box functions.
Evaluations and gradients contaminated with noise.
Very useful tool for finding good model hyper-parameters.
Allows for inequality constraints and several objectives.
Efficient optimization of expensive \textbf{black-box} functions.

- Evaluations and gradients \textit{contaminated with noise}.
- Very useful tool for \textbf{finding good model hyper-parameters}.
- Allows for \textit{inequality constraints} and several objectives.
- **Linear regression model**: Cumulants of the residual above $\kappa_2$ are **reduced in magnitude** in the anti-causal direction.
- A **Gaussianity test** can determine the causal direction.
- The linear assumption is removed by using **kernel methods**.
- **Competitive results** with state-of-the-art methods.

---

**Causal Direction**

- **Fit of the Model**
- **Residuals**

**Anti-Causal Direction**

- **Fit of the Model**
- **Residuals**
Gaussian Process Classification:
- MNIST (60,000 instances): 2.08% (error) and 0.0725 (NLL).
- Airline delays (2 million instances)
- Advantages of the predictive distribution: active learning, outlier detection, etc.

Deep Gaussian Processes Regression:
- Predicting the efficiency of organic photovoltaic molecules.
- 50,000 data instances, 512 dimensions.
- DGPs with 5 layers outperform other models in terms of NLL.

Bayesian Optimization:
- Avoid local minima in hyper-parameter optimization.
- Find neural networks that are both accurate and fast.
- Find classification ensembles that are small and accurate.
- Tune Hamilton Monte Carlo enforcing convergence.

Non-linear Causal Inference:
- Experiments on the ChaLearn cause effect pairs.


Collaborations with Instituto de Ingeniería del Conocimiento (wind power forecasting).

Collaborations with other universities from Madrid under project CASICAM-CM (approximate inference and adversarial learning).

Collaborations with other universities from Madrid under the DAMA network (Bayesian optimization of the parameters of a system for wave energy prediction).

Networks: DAMA (TIN2015-70308-REDT). Diversificación Avanzada de Máquinas de Aprendizaje. Bayesian optimization and Bayesian methods.

Personal Interest and Contributions to KERMES

■ Personal interests:
  ■ New methods for efficient inference with Gaussian processes.
  ■ New methods for Bayesian optimization.
  ■ New methods for approximate inference with Deep GPs.
  ■ Practical applications for all the methods described.

■ What can I bring:
  ■ Extensive knowledge about all the techniques described.
  ■ Experience with efficient coding and implementation of these or new methods (Theano and or Tensorflow).
  ■ Source code for all the methods described.
  ■ Some of my workforce and of the students supervised by me.
  ■ Computational power of the CCC-UAM (two clusters with 1,000 cores each, two Teslas K-40 and more to come).
Expected Contributions from KERMES

- Find and meet people working on similar topics as me.
- Collaborate on interesting problems and applications!
- Get a few publications in COREA-A,B conferences or JCR indexed journals.