kernelPSI: a Post-Selection Inference Framework for Nonlinear Variable Selection

ICML 2019, Long Beach

joint work with
Chloé-Agathe Azencott, Clément Chatelain and Jean-Philippe Vert

Photo credits: ©

Lotfi SLIM
Mines ParisTech & SANOFI

Post-Selection Inference (PSI)

Why does it matter?

- **PSI: performing statistical inference *after* model selection**

Hypothesis testing *e.g.* significance of the constructed model or a single coefficient

Feature selection

+

Need to *account* for the selection event for *valid* inference!

In *practice*,

- Data splitting: loss of accuracy in selection and power in inference

<<

- Exact PSI: new momentum thanks to the *polyhedral lemma*, several setups covered

Absence of nonlinear effects and interactions among covariates!

kernelPSI: key ideas

A generalization of linear methods

Select a subset of kernels that are most associated with an outcome Y, and then to measure the significance of their association with Y (individually or in a joint manner)

Quadratic kernel association scores

$$s(K, Y) = Y^T Q(K) Y$$

- Prototypes: for Q p.d., $s(K, Y) = \left\|\hat{Y}_K\right\|^2$: kernel PCA and kernel ridge regression
- $s(K, Y) = \widehat{HSIC}\,(K, YY^T)$

Selection

For a quadratic kernel association score, select a subset of kernels by:

- **Filtering**
- **Forward/backward stepwise selection**

Selection event = a conjunction of quadratic constraints

Inference

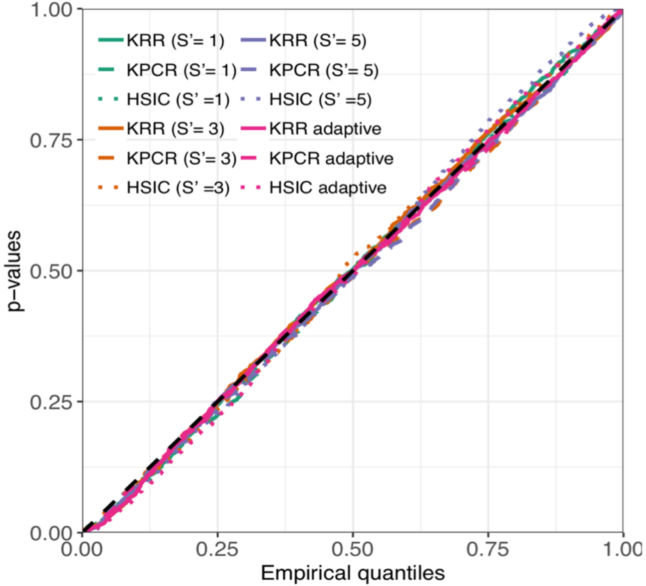
In a model of the form $Y = \mu + \sigma^2\epsilon$, test for a null hypothesis:

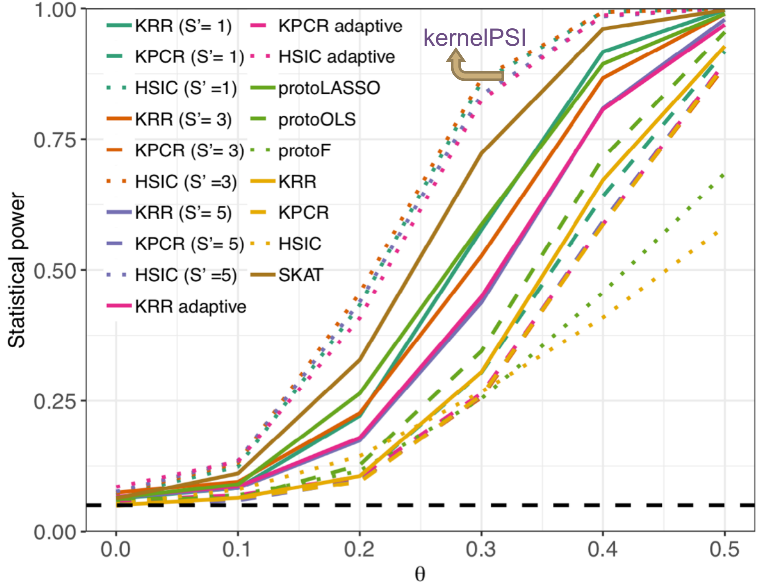
$$H_0: s(K, \mu) = 0$$

- **Test statistics: norm of prototypes, log-likelihoods**
- **Constrained sampling to generate empirical p-values**

SANOFI | MINES ParisTech

kernelPSI: results





The obtained p-values are valid!

KernelPSI is better than non-selective kernel baselines and selective linear ones!

kernelPSI: poster #228

THANK YOU!

Contact details:

lotfi.slim@[mines-paristech.fr, sanofi.com]

SANOFI

MINES ParisTech