Classification, Induction and Brain Decoding

Emanuele Olivetti, Susanne Greiner

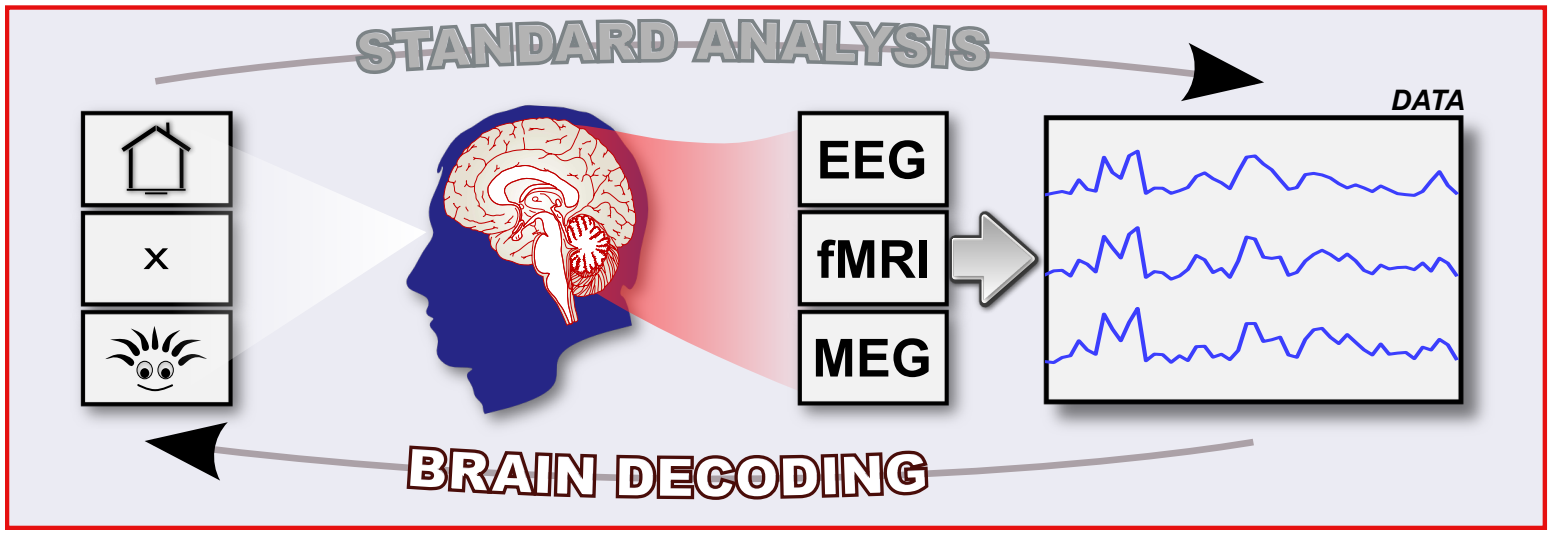
NeuroInformatics Laboratory (NILab)
Bruno Kessler Foundation, Trento (FBK), Italy
Center for Mind and Brain Sciences (CIMeC), University of Trento, Italy
http://nilab.fbk.eu
olivetti@fbk.eu
greiner@fbk.eu

Python in Neuroscience Workshop (EuroSciPy 2011)
Paris, Ecole Normale Supérieure, August 29-30 2011

Motivation: Brain Decoding [Haynes and Rees, 2006]

A topic relevant for BCI and Neuroscience research...

What is *brain decoding*?

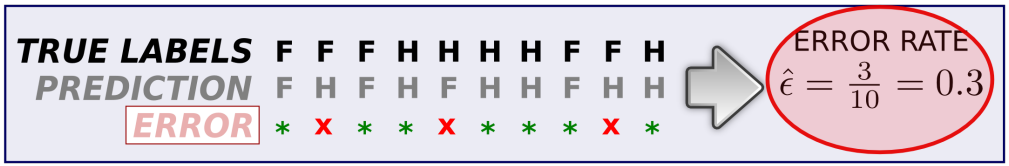


TRUE LABELS	F	F	F	H	H	H	H	F	F	H
PREDICTION	F	H	F	H	F	H	H	F	H	H
ERROR	*	x	*	*	x	*	*	*	x	*

ERROR RATE $\hat{\epsilon} = \frac{3}{10} = 0.3$

Underlying (Inductive) Questions

- "*Is there stimulus-related information within brain data?*" [Pereira et al., 2009]
- Information $\Longleftrightarrow$ Classification error?
- Is *my* classifier better than the *random* classifier?
- Did the classifier learn a thing?
- How to make inferences from multi-subjects experiments?



Does this really matter? YES

- Datasets are small and very high-dimensional. It is difficult to have high confidence in the answers.

Small Sample Mythology

- The (estimated) error rate is a good performance measure.

A:

	F	H
F	90	10
H	0	0

$\hat{\epsilon} = \frac{10}{100}$

B:

	F	H
F	80	0
H	10	10

$\hat{\epsilon} = \frac{10}{100}$

columns: TRUE CLASS LABELS
rows: PREDICTED CLASS LABELS

- The variance of the cross-validation error rate.
(`http://hunch.net/?p=29`)
[Bengio and Grandvalet, 2004]
- The error rate is Gaussian distributed.
- When doing multi-subject inferences the within-subject variance can be neglected.

Our Answers

1. Information $\Longleftrightarrow$ Classification error?
   [Olivetti et al., 2011a]
2. Did the classifier learn a thing?
   [Current Work]
3. How to make inferences from multi-subjects experiments?
   [Olivetti et al., 2011b]

1. Information $\Leftrightarrow$ good classif.? [Olivetti et al., 2011a]

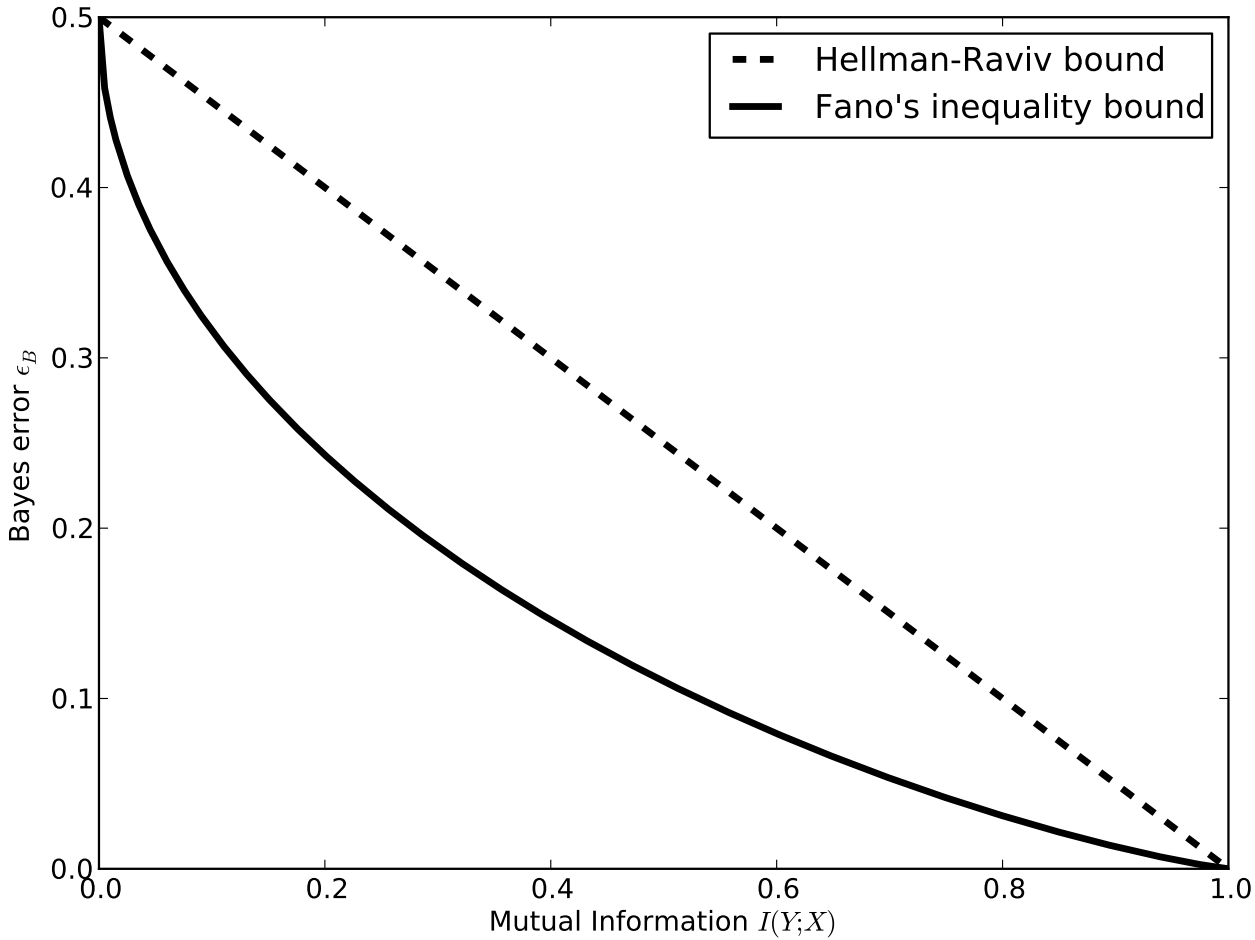


Figure: Plot of the Fano's and Hellman-Raviv's inequalities in terms of mutual information and Bayes error *for the binary balanced case*.

2. Did the classifier learn a thing? [Current Work]

- All tests in the literature are based on the *error rate*.
- Is the error rate an appropriate measure? Examples:

A:

	F	H
F	90	10
H	0	0

$\hat{\epsilon} = \frac{10}{100}$

B:

	F	H
F	80	0
H	10	10

$\hat{\epsilon} = \frac{10}{100}$

- Our approach: test of independence of predicted vs. true class-labels.
  - A: $p(\text{independent}|\text{data}) = 76\%$
  - B: $p(\text{independent}|\text{data}) = 0.00015\%$

Claim

accurate inference from small samples requires something more than the error rate, even when classes are not strongly unbalanced.

3. Inductive reasoning about the population

"*Is there information at the population level?*"

Standard Procedure

1. The error rate Gaussian.
2. No *within*-subject variance.
3. *t*-test.
4. Classical hyp. testing.
5. p-value $\leq 0.05 \Rightarrow$ inform.

[Olivetti et al., 2011b]

1. No Gauss. approx. $\rightarrow$ Beta.
2. Account for $\hat{\epsilon}_i$ uncertainty.
3. Hierarchical model.
4. Bayesian hyp. testing.
5. Bayes factor $\geq 3 \Rightarrow$ inform.

Python, NumPy, SciPy...

- Broadcasting/Vectorize magic in NumPy: “Broadcasting provides a means of vectorizing array operations so that looping occurs in C instead of Python.”
- Sampling from a hierarchical model [Olivetti et al., 2011b]: “*15 subjects, test set of size 50, 100k iterations: 1.67 sec.*” See `cool_broadcasting_vectorize.py`
- Implementation of our algorithms is freely available:
  `github.com/emanuele/information_test`
  `github.com/emanuele/Bayes-factor-multi-subject`
  ...
- Of course we are big fans (and sometimes contributors) of PyMVPA, scikits.learn, PyMC, NiPy (Dipy, nibabel,...)!
- We hope to send pull requests soon :-)

Thank You!

slides available at:

`https://github.com/emanuele/pin2011`

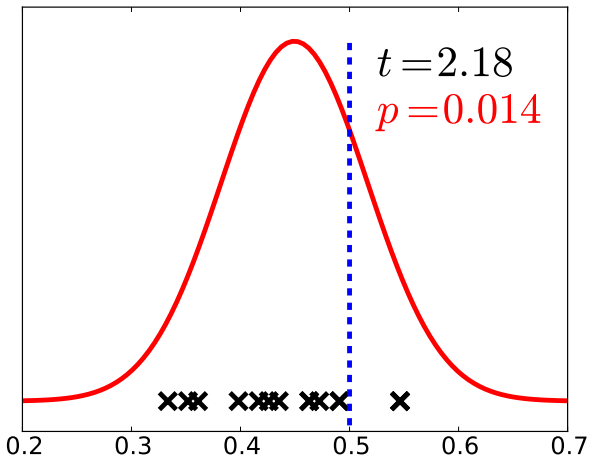
Inductive reasoning about population

"*Is there information at the population level?*"

Ex.: predict mental calculation from PSPL [Knops et al., 2009]

- test set size: 108 presentations $\times$ 15 subjects.
- $\{43, 59, 51, 38, 39, 53, 47, 50, 50, 59, 59, 45, 36, 46, 53\}$.

t-test



[Olivetti et al., 2011b]

- No Gauss. approx. $\rightarrow$ Beta.
- Account for $\hat{\epsilon}_i$ uncertainty.
- Hierarchical model.
- Bayesian hyp. testing.
- $B_{\text{Inf/NoInf}} = 2.5$ (weak)


Bengio, Y. and Grandvalet, Y. (2004).
No Unbiased Estimator of the Variance of K-Fold Cross-Validation.
J. Mach. Learn. Res., 5:1089–1105.


Haynes, J.-D. and Rees, G. (2006).
Decoding mental states from brain activity in humans.
Nature Reviews Neuroscience, 7(7):523–534.


Knops, A., Thirion, B., Hubbard, E. M., Michel, V., and Dehaene, S. (2009).
Recruitment of an Area Involved in Eye Movements During Mental Arithmetic.
Science, 324(5934):1583–1585.


Olivetti, E., Veeramachaneni, S., and Avesani, P. (2011a).
Testing for Information with Brain Decoding.
In *IEEE International Workshop on Pattern Recognition in NeuroImaging*.


Olivetti, E., Veeramachaneni, S., and Nowakowska, E. (2011b).
Bayesian hypothesis testing for pattern discrimination in brain decoding.
Pattern Recognition.


Pereira, F., Mitchell, T., and Botvinick, M. (2009).
Machine learning classifiers and fMRI: A tutorial overview.
NeuroImage, 45(1):199–209.