Multivariate pattern analysis in brain imaging

Bertrand Thirion

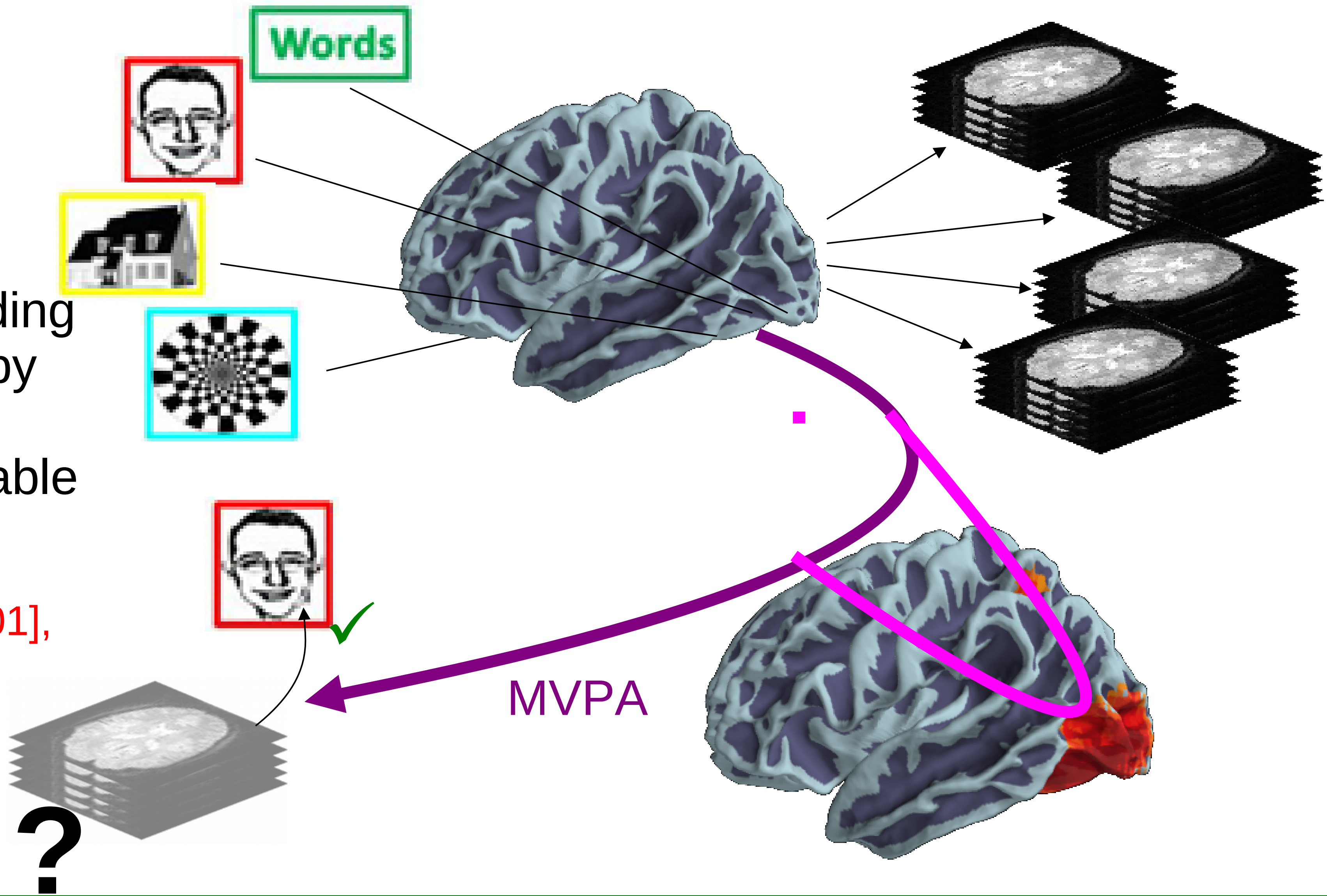
bertrand.thirion@inria.fr

PARIETAL

Basic concepts

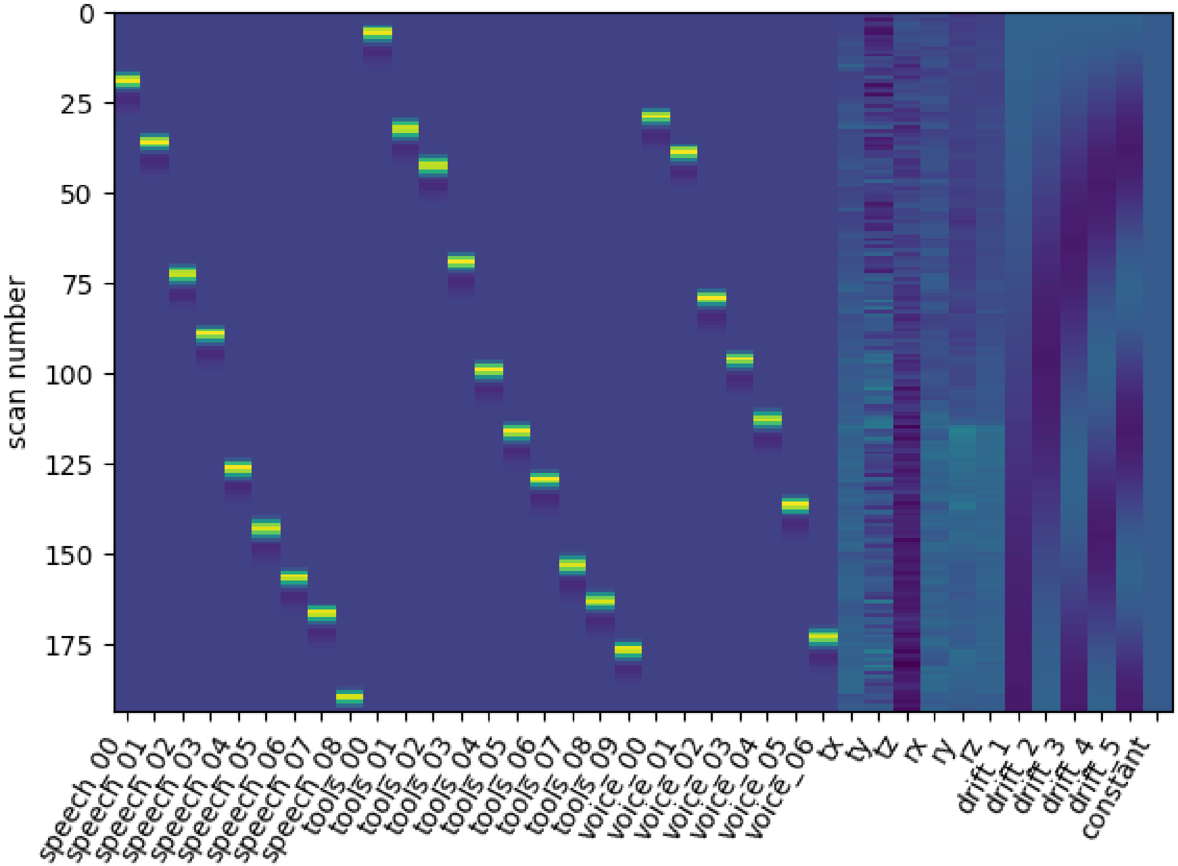
Multivariate pattern analysis

Aims at decoding brain activity by predicting a cognitive variable
[Dehaene et al. 1998], [Haxby et al. 2001], [Cox et al. 2003]



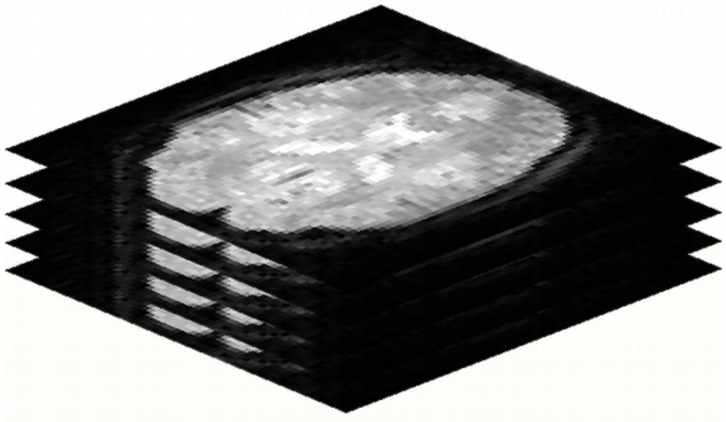
A pipeline view

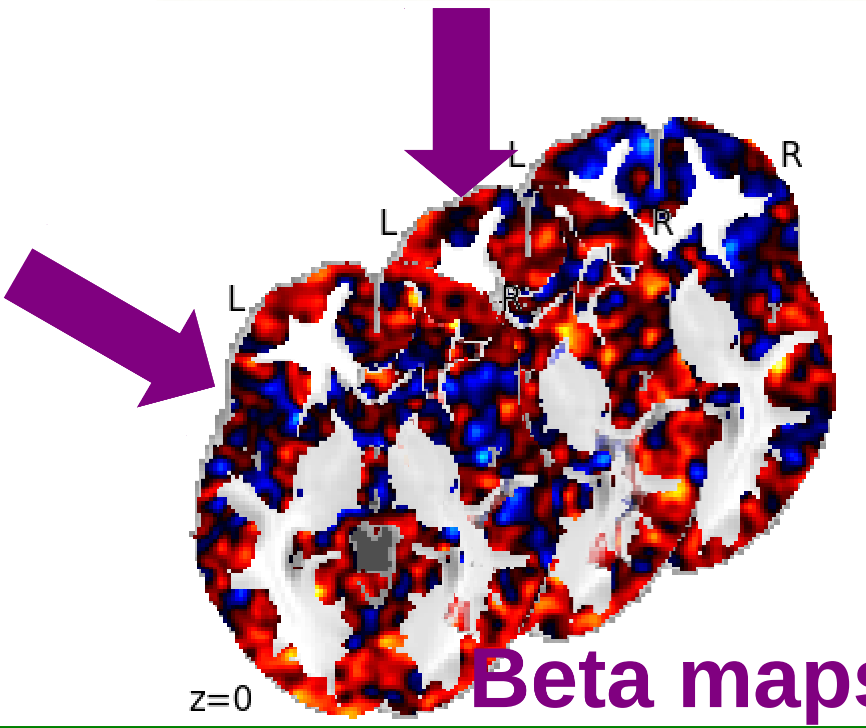
Experimental events
→ trial-wise design matrix



Each event belongs to a class

FMRI data

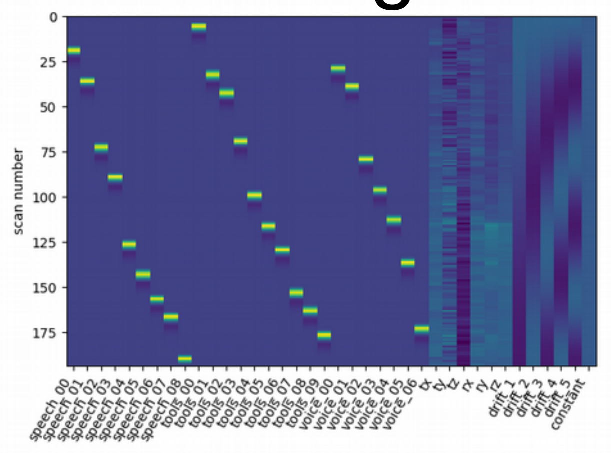




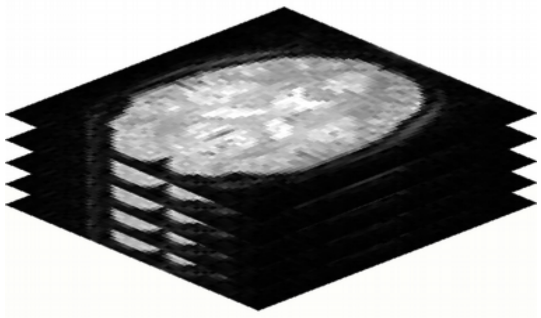
Beta maps

A pipeline view

trial-wise design matrix



FMRI data



Events class

music

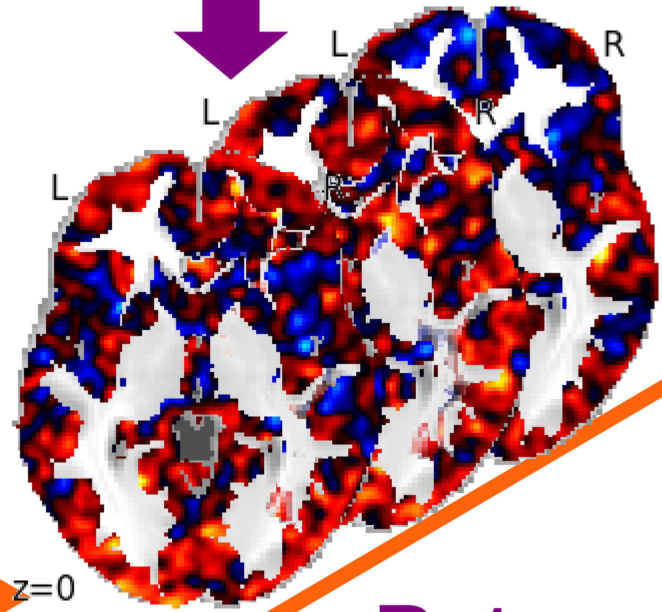
...

speech

music

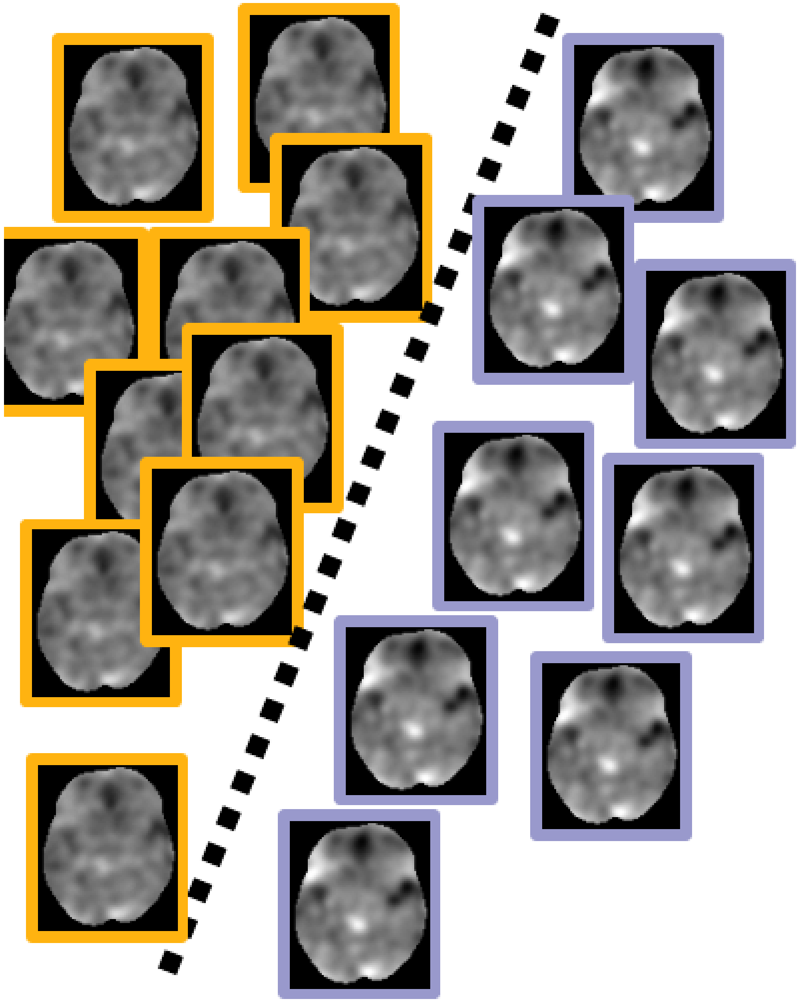
...

Decoding model



Beta maps

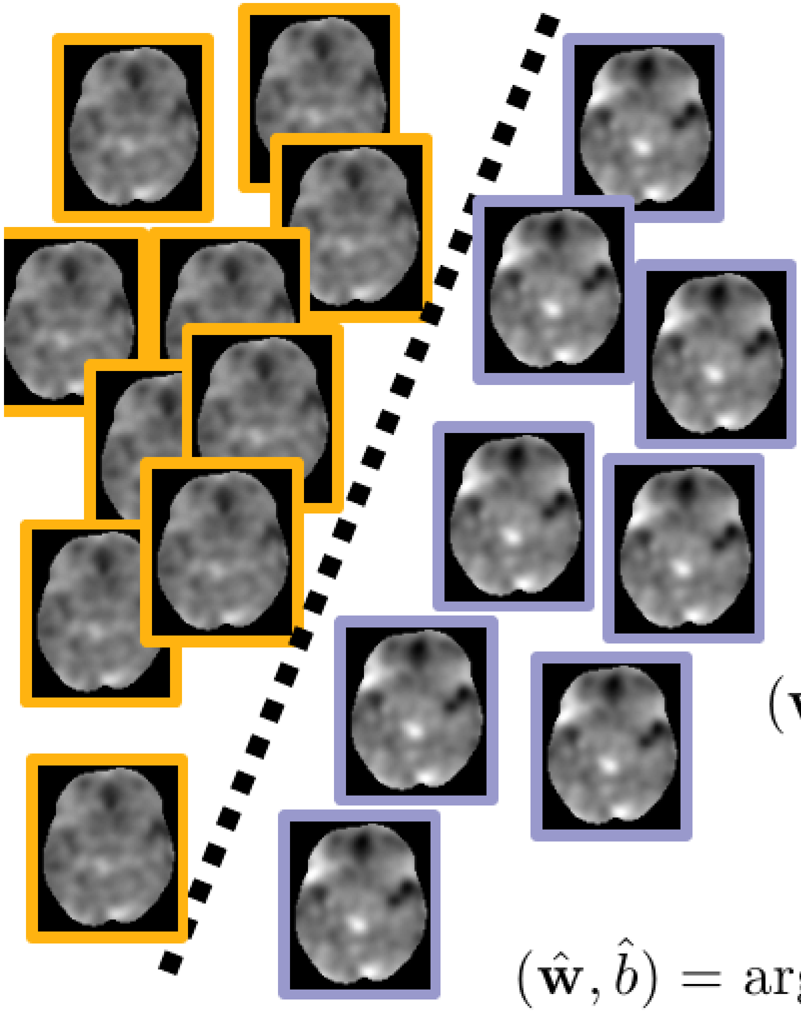
Image-based classification



- Given x in R^p, (*MRI volume with p voxels), predict a label y in {-1, 1} i.e. ▮ or ▮

  or better the class probability Proba(y = 1|x)

Image-based classification



- Given x in R^p, (*MRI volume with p voxels), predict a label y in {-1, 1} i.e. ■ or ■

  or better the class probability Proba(y = 1|x)

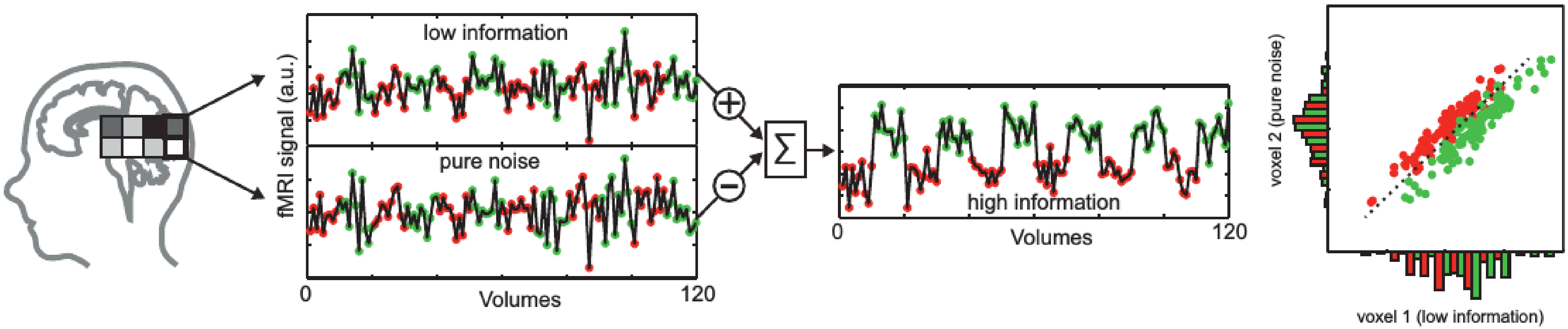
- Use of logistic regression: learn the weight w and bias b such that

$$(\hat{\mathbf{w}}, \hat{b}) = \mathrm{argmin}_{\mathbf{w},b} \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i(\mathbf{x}_i^T\mathbf{w} + b)\right)\right)$$

- With regularization

$$(\hat{\mathbf{w}}, \hat{b}) = \mathrm{argmin}_{\mathbf{w},b} \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i(\mathbf{x}_i^T\mathbf{w} + b)\right)\right) + \lambda\|\mathbf{w}\|_2^2$$

The dream case for MVPA



- Individual voxels corrupted by a noise source → weakly significant
- Their difference is strongly task related: accurate classification

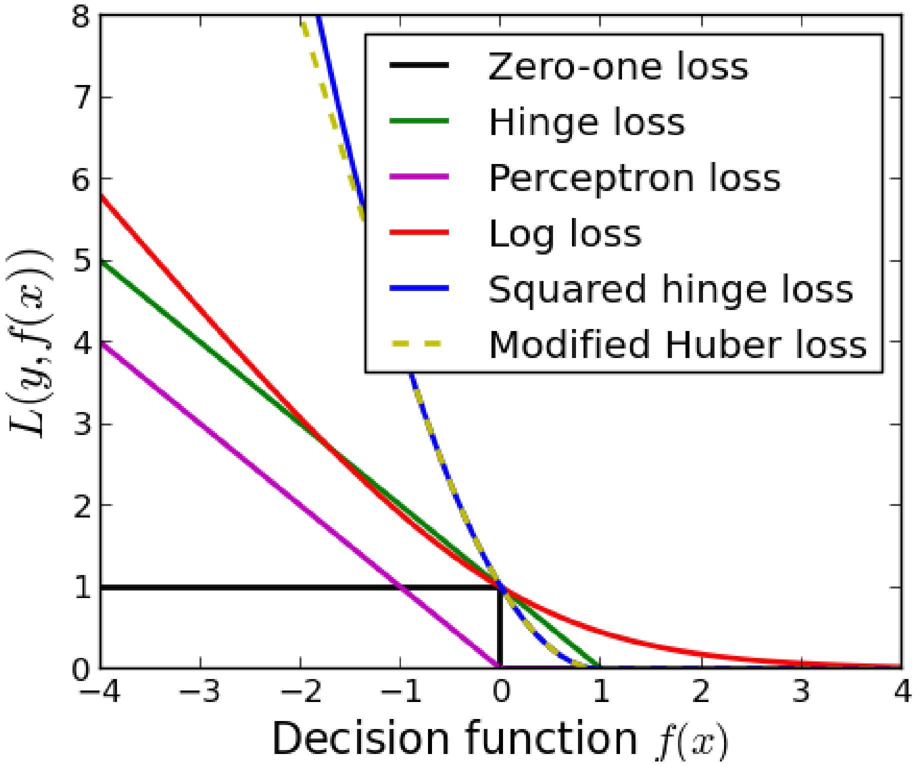
[Haufe et al. nimg 2013, Haynes neuron 2015]

Training a predictive model

- **Learning pattern w from training set (y, X)**

$$\hat{\mathbf{w}} = \mathrm{argmin}_{\mathbf{w}\in\mathbb{R}^p} \sum_{i=1}^{n} \ell(\mathbf{y_i}, \mathbf{X_i}\mathbf{w}) + \lambda J(\mathbf{w})$$

- **Choice of the loss**
  - Regression: Least-squares, Hinge, Huber
  - Classification: Hinge, logistic
- **Choice of the regularizer**
  - Predictive setting: a norm on **w**
  - Bayesian setting: prior distribution on **w**

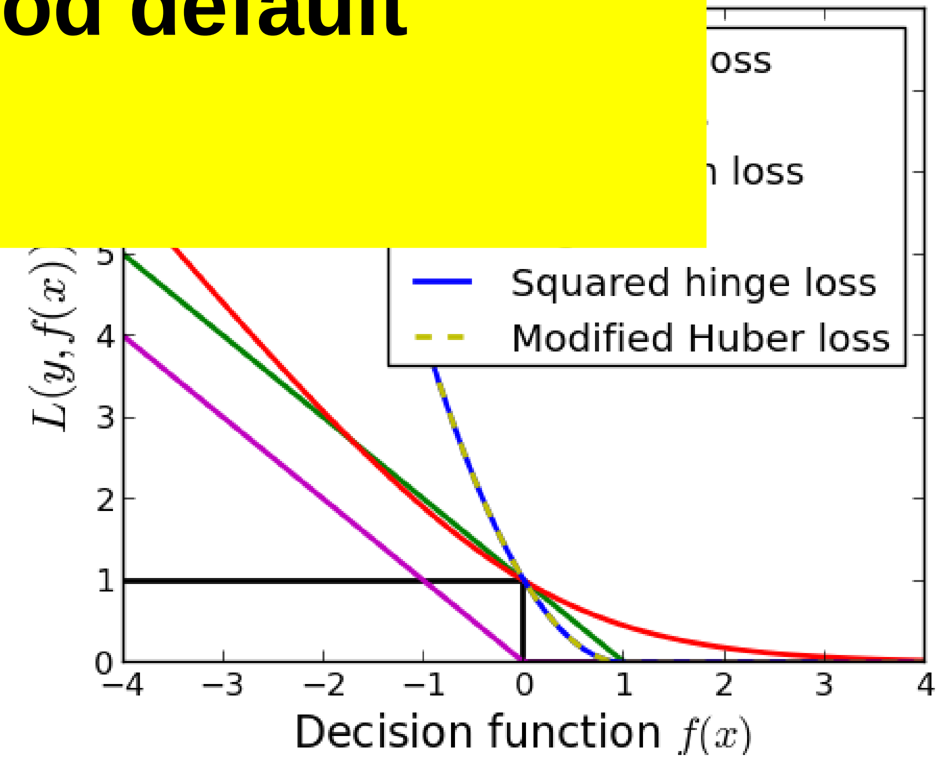


Training a predictive model

- **Learning pattern w from training set (y, X)**
- **Choice o**
  - Regre
    Hinge
  - Class
- **Choice of the regularizer**
  - Predictive setting: a norm on **w**
  - Bayesian setting: prior distribution on **w**

$$\sum^{n} \ldots + \lambda J(\mathbf{w})$$

SVM classification/regression (C=1) is a good default



Evaluation of the decoding

Measure prediction accuracy

https://scikit-learn.org/stable/modules/model_evaluation.html

Regression →
Explained variance ζ :

$$\zeta(\mathbf{y^t}, \hat{\mathbf{y}}) = \frac{\mathrm{var}(\mathbf{y^t}) - \mathrm{var}(\mathbf{y^t} - \hat{\mathbf{y}})}{\mathrm{var}(\mathbf{y^t})}$$

Classification score:

$$\kappa(\mathbf{y^t}, \hat{\mathbf{y}}^\mathbf{t}) = \frac{\sum_{i=1}^{n^t} \delta(y_i^t, \hat{y}_i^t)}{n^t}$$

→ amount of information about y in the brain data

Cross validation

```
X = np.random.randn(*fmri_masked.shape) # replace with null data
prediction = svc.fit(X, conditions).predict(X)
print((prediction == conditions).sum() / float(len(conditions)))
```

```
1.0
```

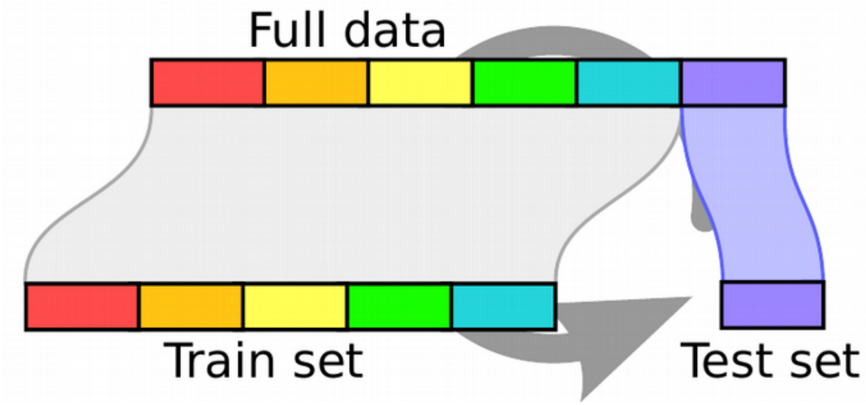
If test data = training data, you get 100% accuracy, even when your data are noise

Cross validation

```
X = np.random.randn(*fmri_masked.shape) # replace with null data
prediction = svc.fit(X, conditions).predict(X)
print((prediction == conditions).sum() / float(len(conditions)))
```

```
1.0
```

If test data = training data, you get 100% accuracy, even when your data are noise



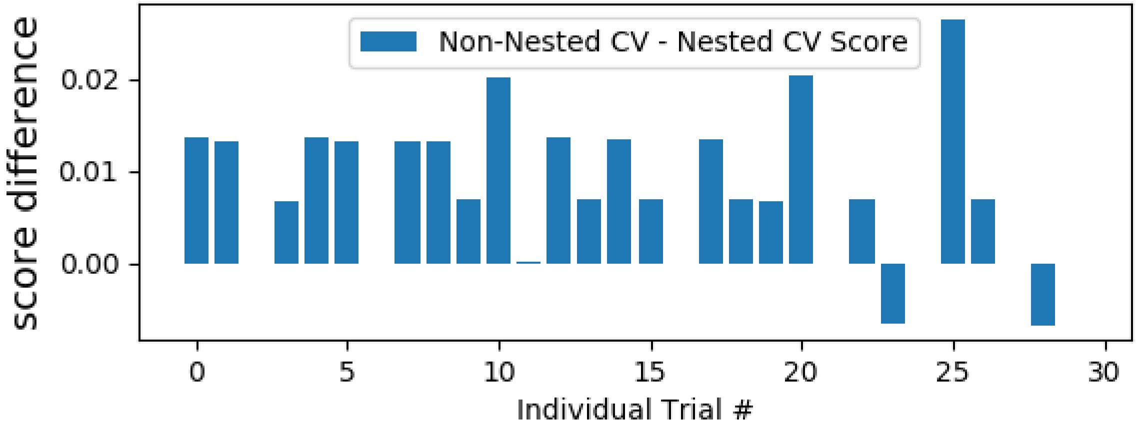
With cross-validation, accuracy is unbiased

```
cv_score = cross_val_score(svc,
                           X,
                           conditions,
                           cv=cv, # cross-validation scheme
                           groups=session_label, # data splitting scheme
                           )
print(np.mean(cv_score)) # chance is 50%
```

```
0.4212962962962962
```

Parameters to tune

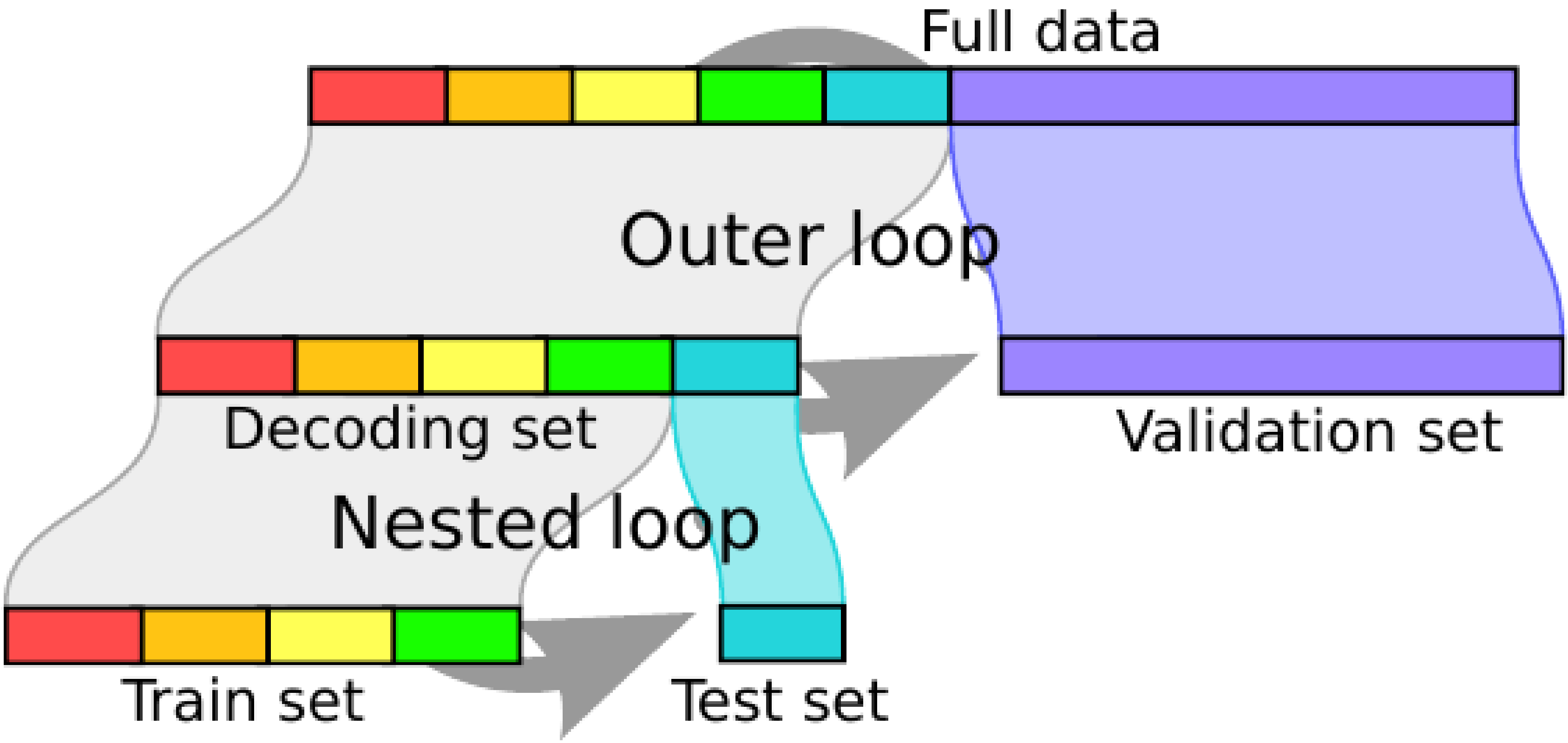
- Parameter/model selection → based on accuracy
- CAVEAT: do not do it with the scoring loop



https://scikit-learn.org/stable/auto_examples/model_selection/plot_nested_cross_validation_iris.htm

- Need nested loop **“nested cross validation”**

Nested cross-validation

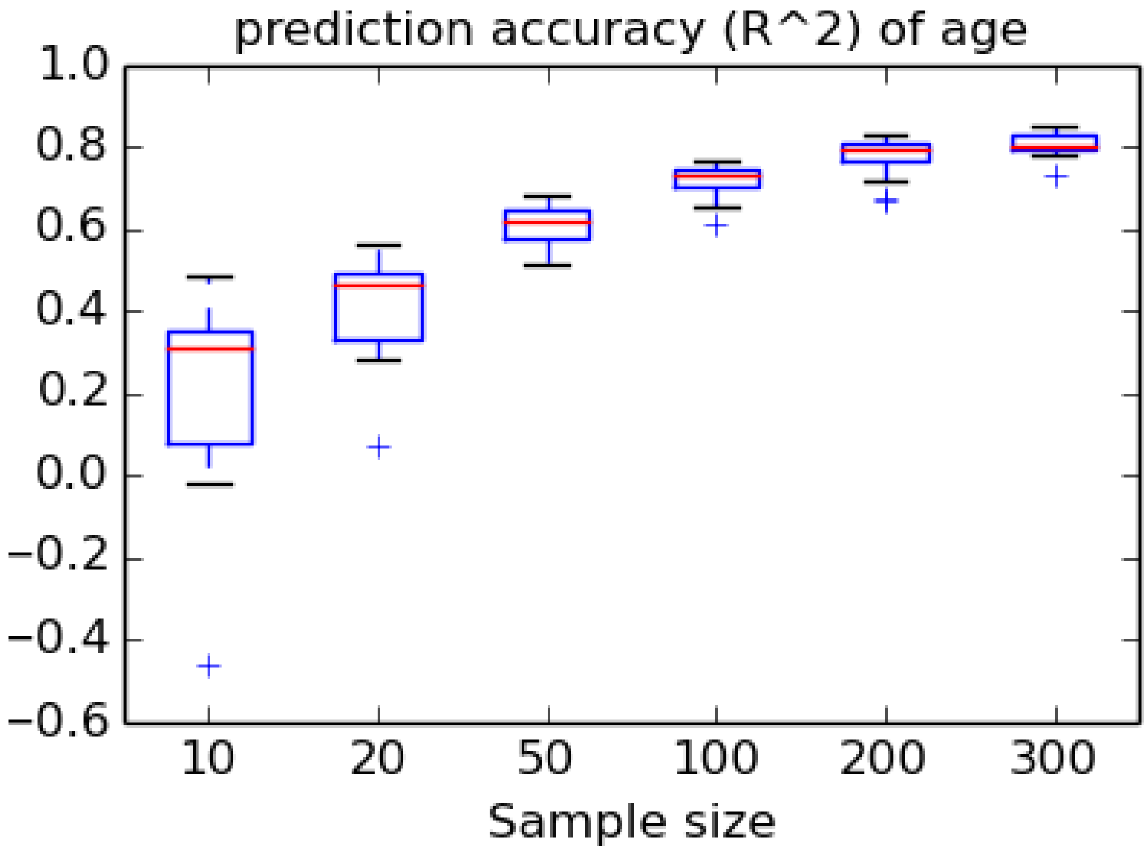


- One loop to tune inner parameters
- One loop to get the accuracy

Common pitfalls

Learning curve: how prediction improves with n

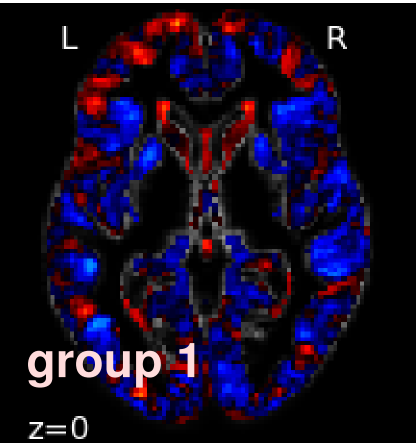
- Predict the age of a subject given gray matter density maps (OASIS dataset, n=403)



Weight maps for age prediction / OASIS

The weight map depends on the batch of subject considered (bootstrap):

One question, different datasets, different answers



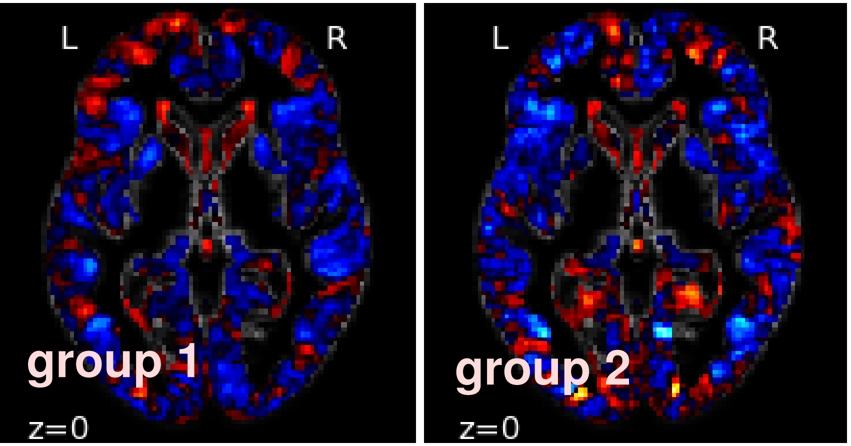
group 6

Variability actually worse than for univariate analysis !

Weight maps for age prediction / OASIS

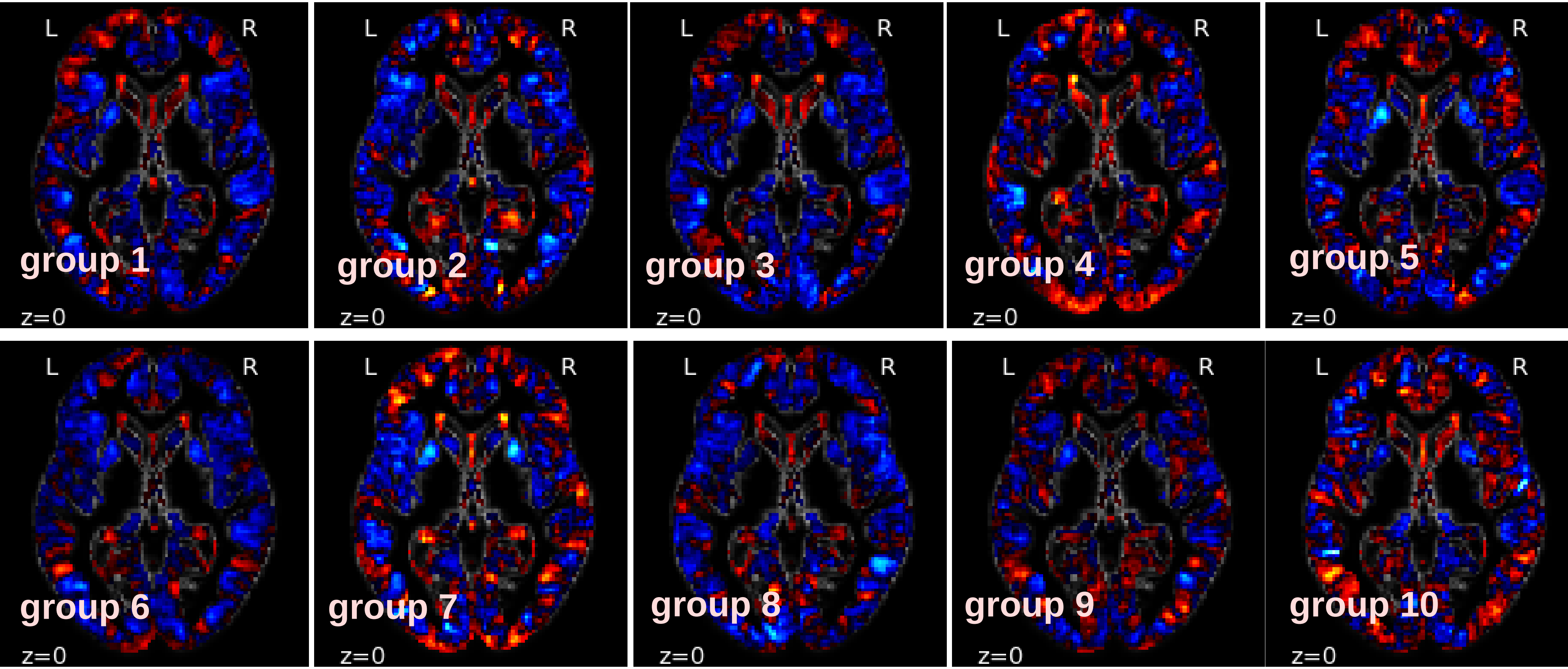
The weight map depends on the batch of subject considered (bootstrap):

One question, different datasets, different answers



Weight maps for age prediction / OASIS

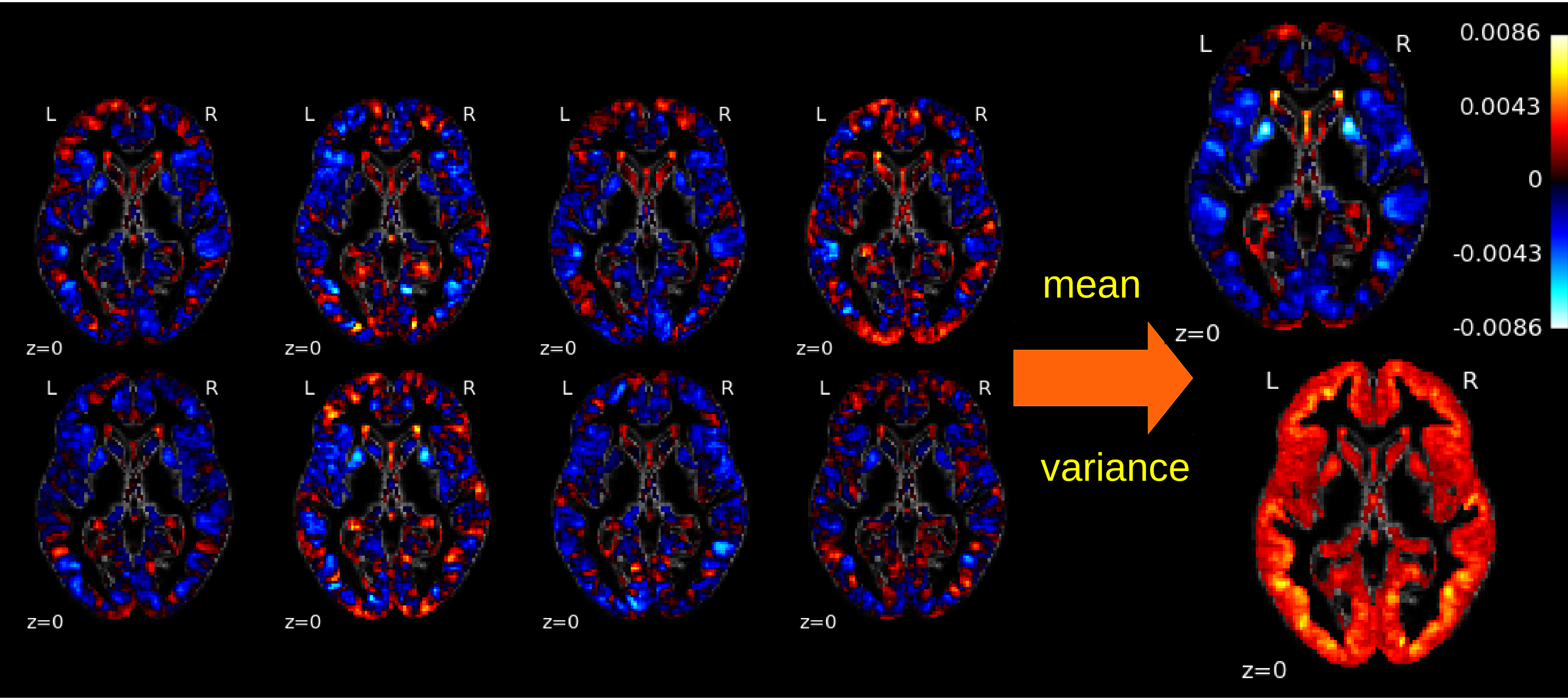
The weight map depends on the batch of subject considered (bootstrap):
One question, different datasets, different answers



Variability actually worse than for univariate analysis !

Weight maps for age prediction / OASIS

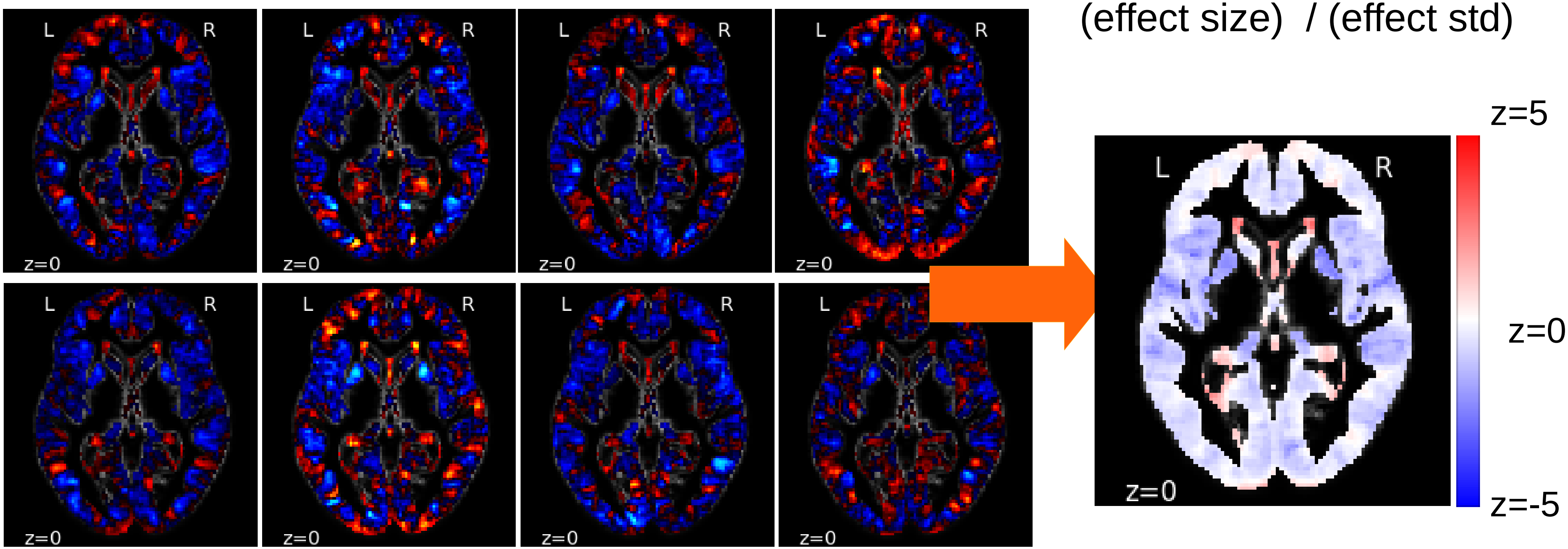
The weight map depends on the batch of subject considered (bootstrap):
One question, different dataset, different answers



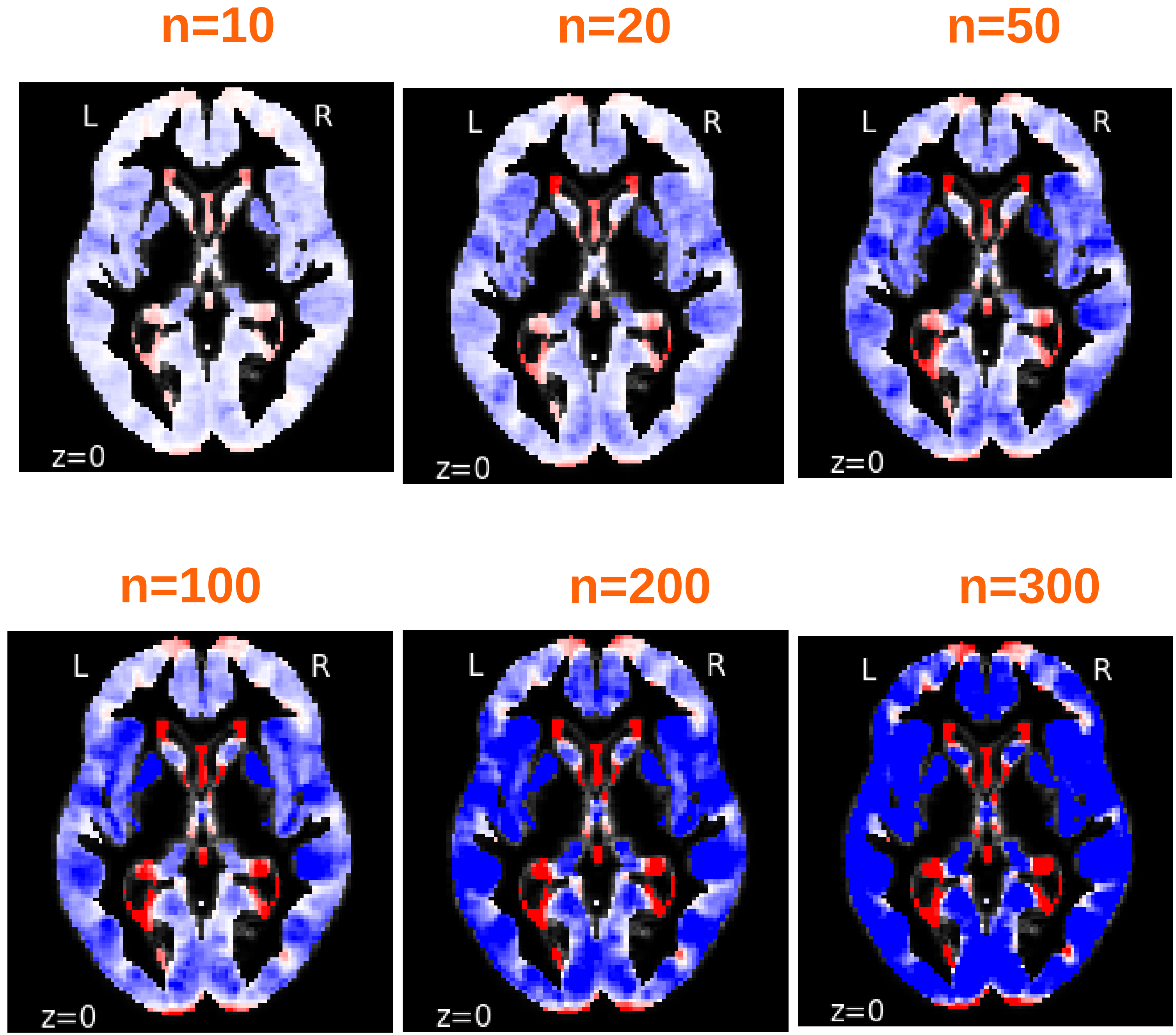
Weight maps for age prediction / OASIS

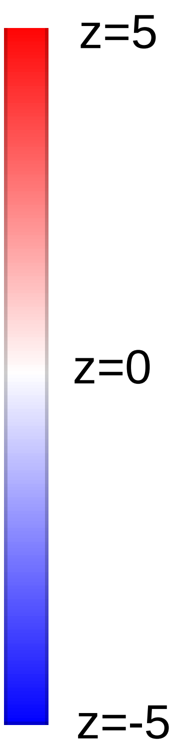
The weight map depends on the batch of subject considered (bootstrap):
One question, different dataset, different answers

Summarized into a z image:
(effect size) / (effect std)



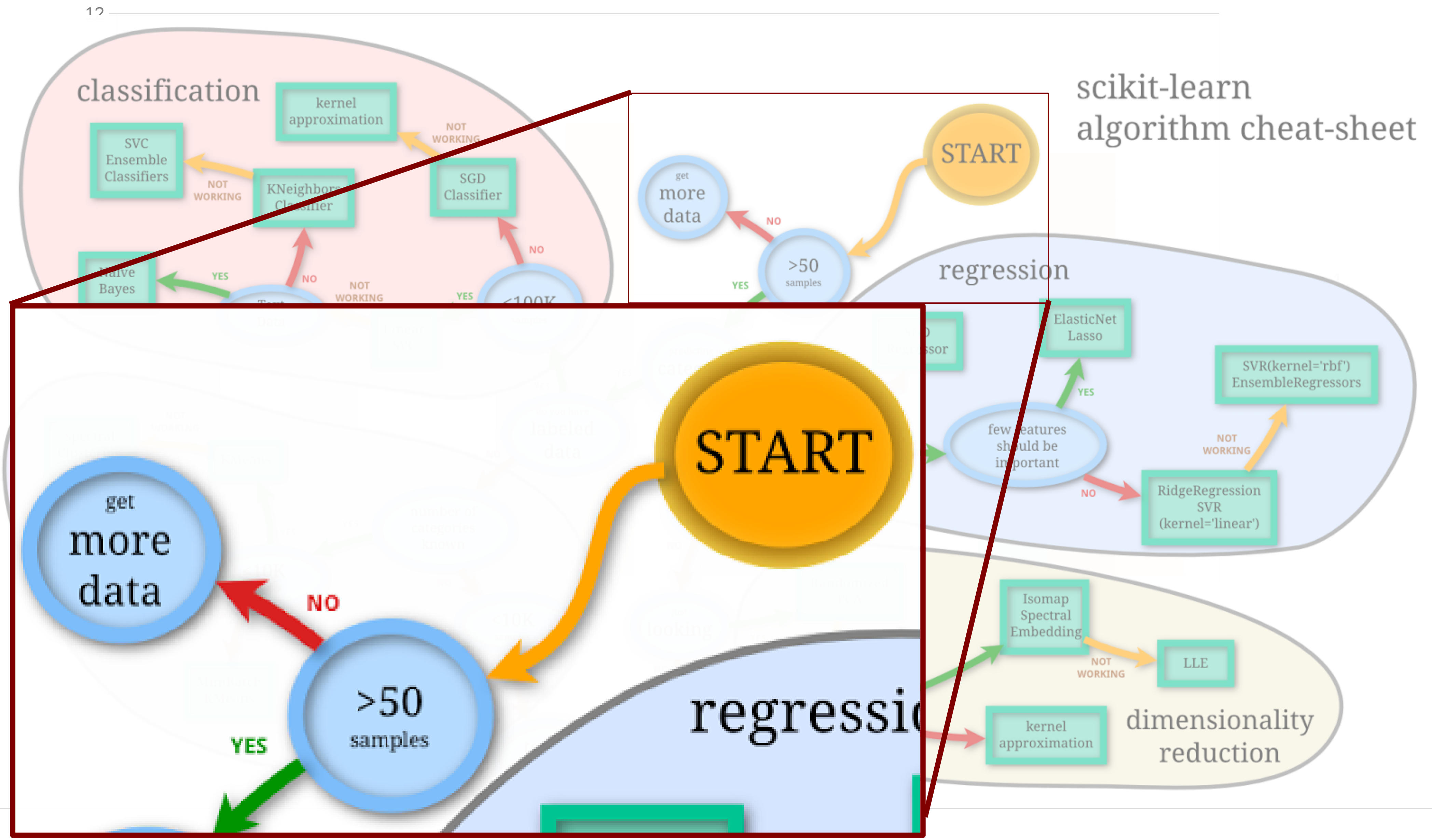
Weight maps for age prediction / OASIS



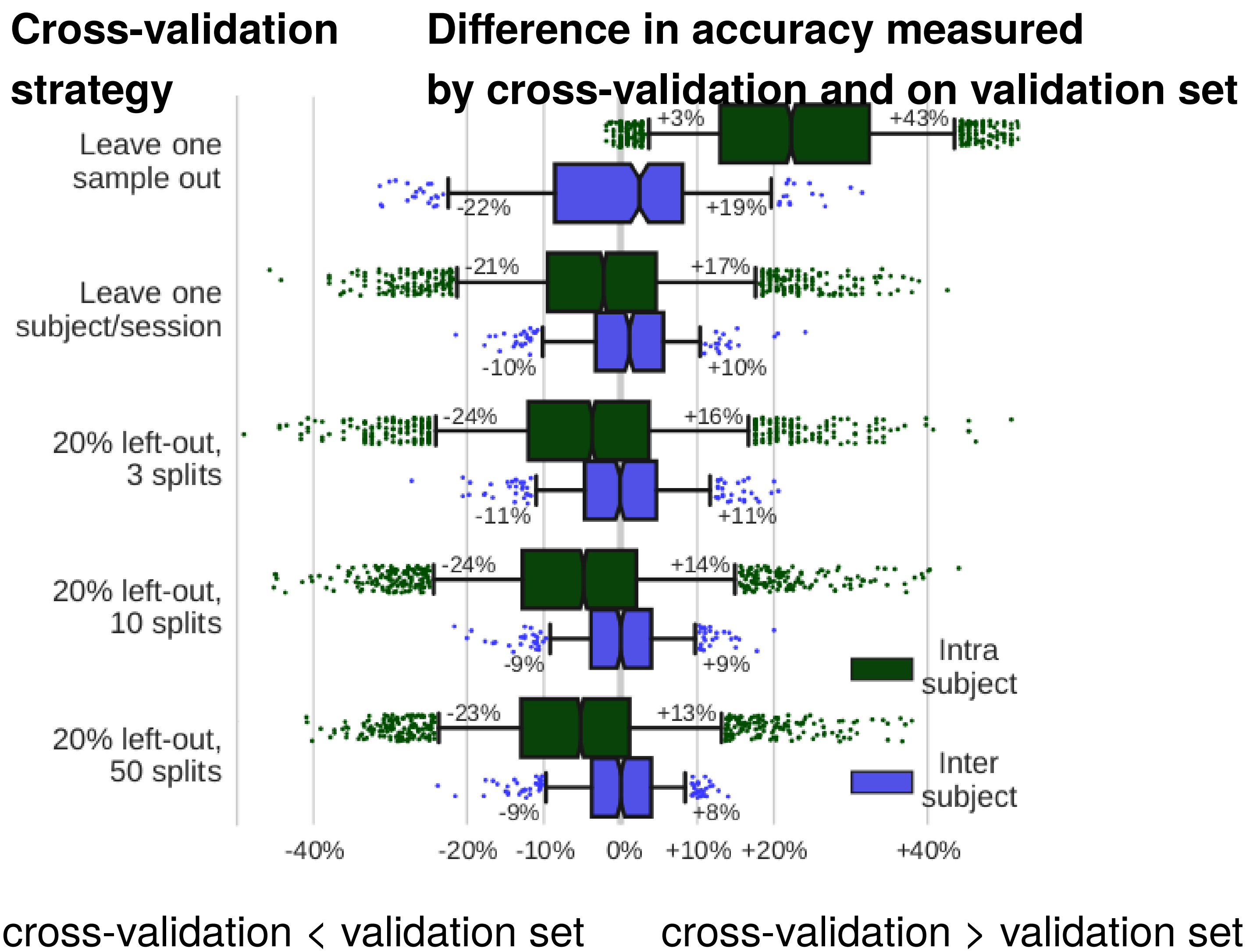


(effect size estimated by bootstrap)

Multivariate analysis



Sample size & cross-validation



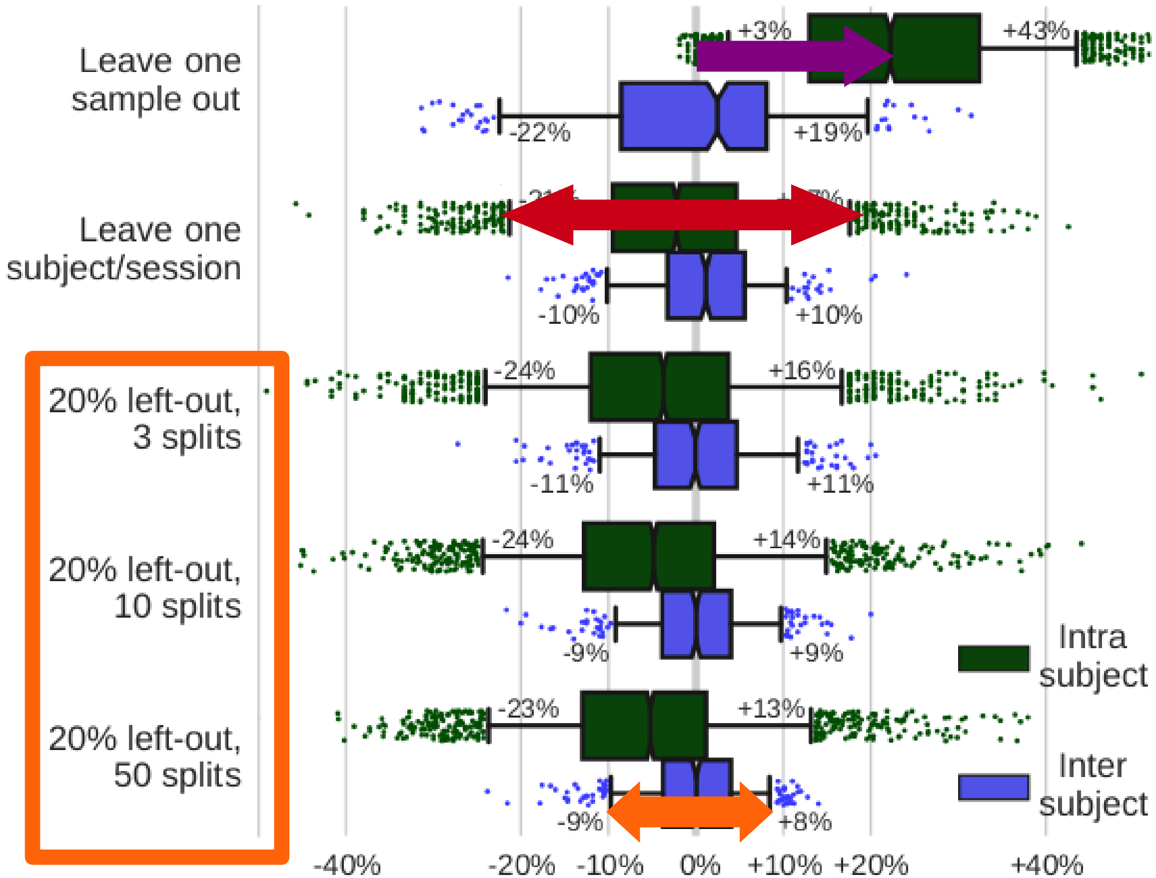
Large-scale experiment:
4 classifiers, 7 datasets, 1 anatomical dataset, 1 MEG dataset

[Varoquaux et al. NIMG 2016]

Sample size & cross-validation

Cross-validation strategy

Difference in accuracy measured by cross-validation and on validation set



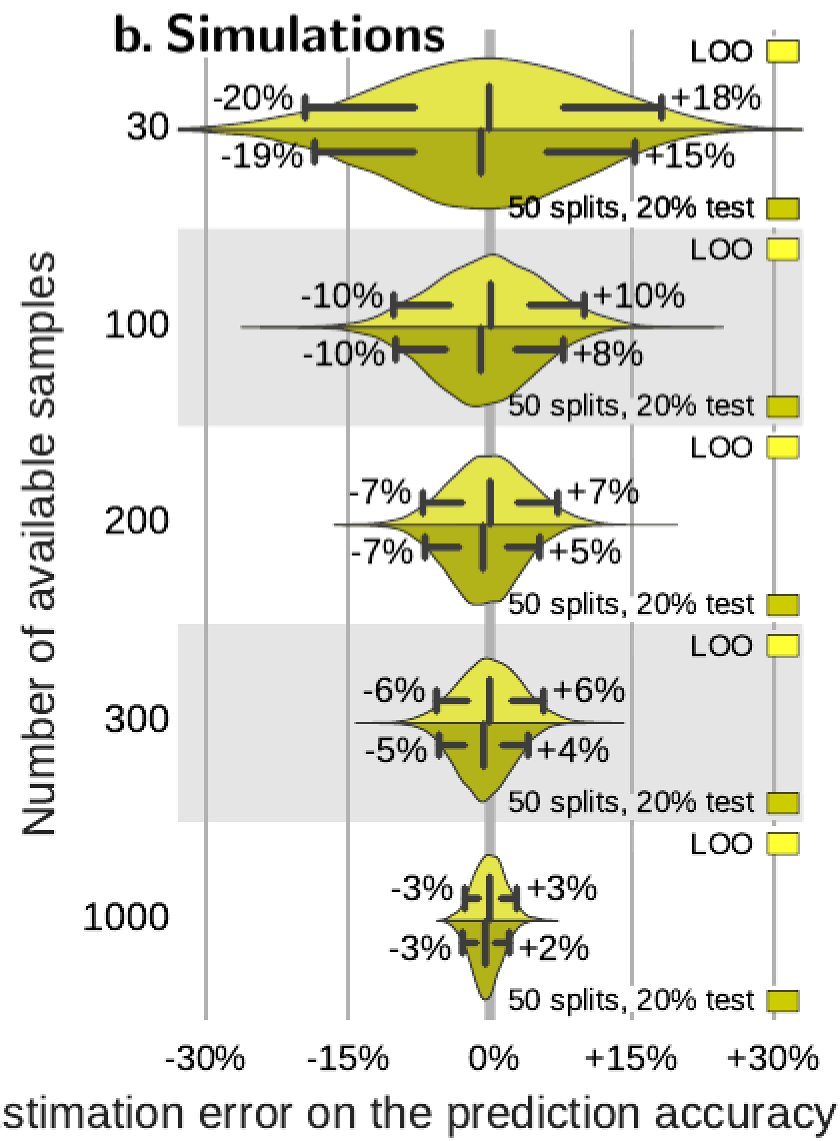
cross-validation < validation set   cross-validation > validation set

- optimistic bias in LOO for non-independent samples
- higher variance in LOO
- variance large overall

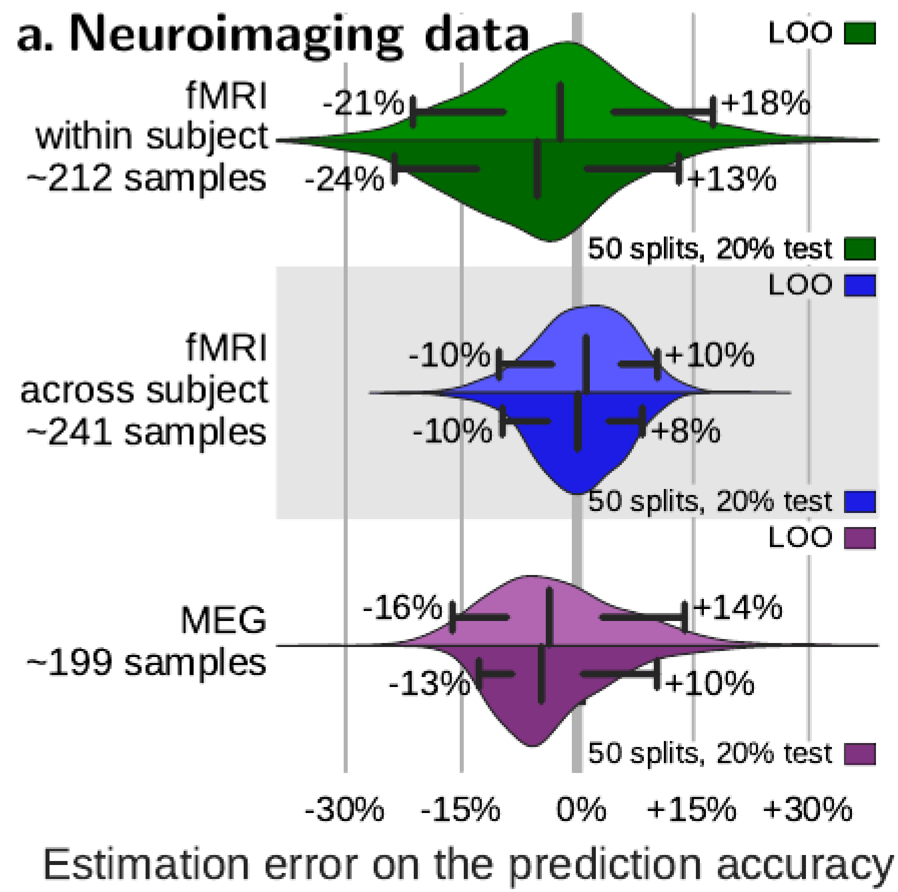
→ use shuffle-split with many splits

[Varoquaux et al. NIMG 2016]

Sample size and cross-validation



Rule of the thumb: uncertainty in prediction decreases with 1/ √n



[G. Varoquaux nimg 2017]

Why don’t we use deep neural networks ?

- They are expensive, hard to tune, hard to interpret and do not bring performance gains in most neuroimaging settings

https://www.biorxiv.org/content/10.1101/473603v1

Do it yourself !

http://nilearn.github.io/