Machine Learning applied to Planetary Sciences

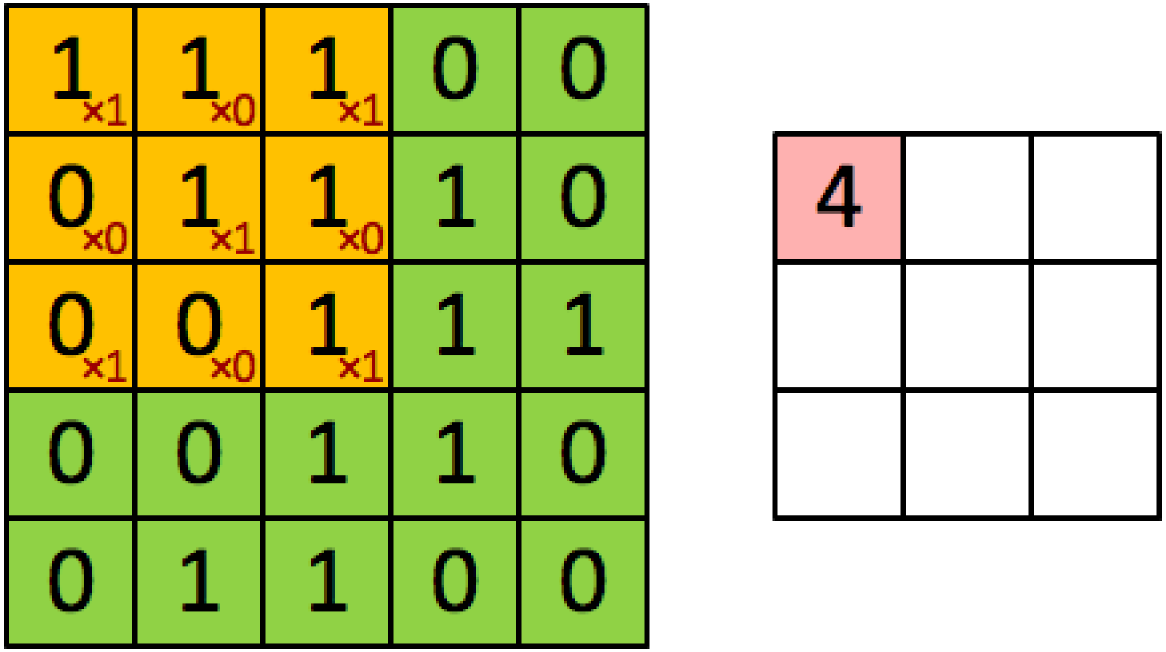
PTYS 595B/495B

Leon Palafox

https://leonpalafox.github.io/MLClass/

Convolutional Neural Nets

Convolution



$1_{\times 1}$	$1_{\times 0}$	$1_{\times 1}$	0	0
$0_{\times 0}$	$1_{\times 1}$	$1_{\times 0}$	1	0
$0_{\times 1}$	$0_{\times 0}$	$1_{\times 1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

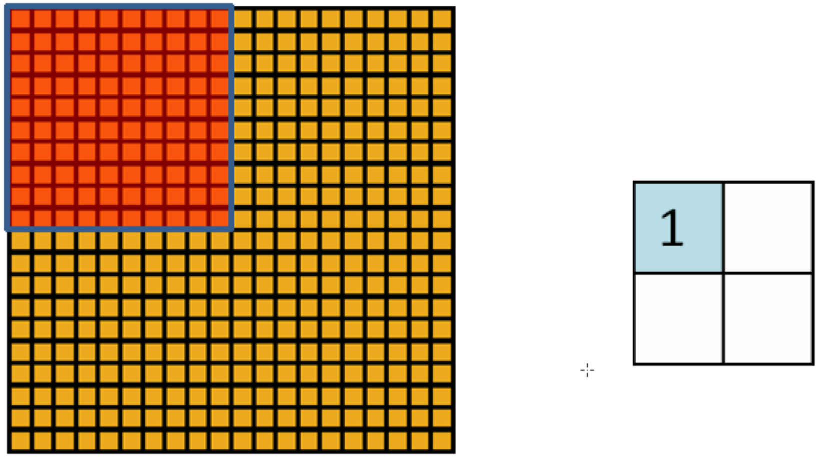
4		

Convolved Feature

Pooling

- Once we have learned the convolved features, we need to take advantage of the locality.
- We choose adjacent features, and can either take the max or the mean .
- The size of the pooling is defined by the user.
- This way we reduce the number of features and at the same time we take advantage of locality.

Pooling



Convolved feature

Pooled feature

Analysis

- By the end of the training a CNN training scheme is similar to training with an artificially large dataset.
  - Similar results
- Pooling actually decreases the number of weights in the actual network (The autoencoder did most of the heavy lifting)
- Sharing weights is the reason the CNN takes into account local features instead of global ones.

Disadvantages

- This approach is ad-hoc for images (or look alike).
- Trying to use it in time-series or other 1D data is not necessarily a good idea.
  - Long training times
- Unless you use Theano/TensorFlow/MatConvNet/Torch is hard to do real work.

Validation Methods

This is were we know who is worthy

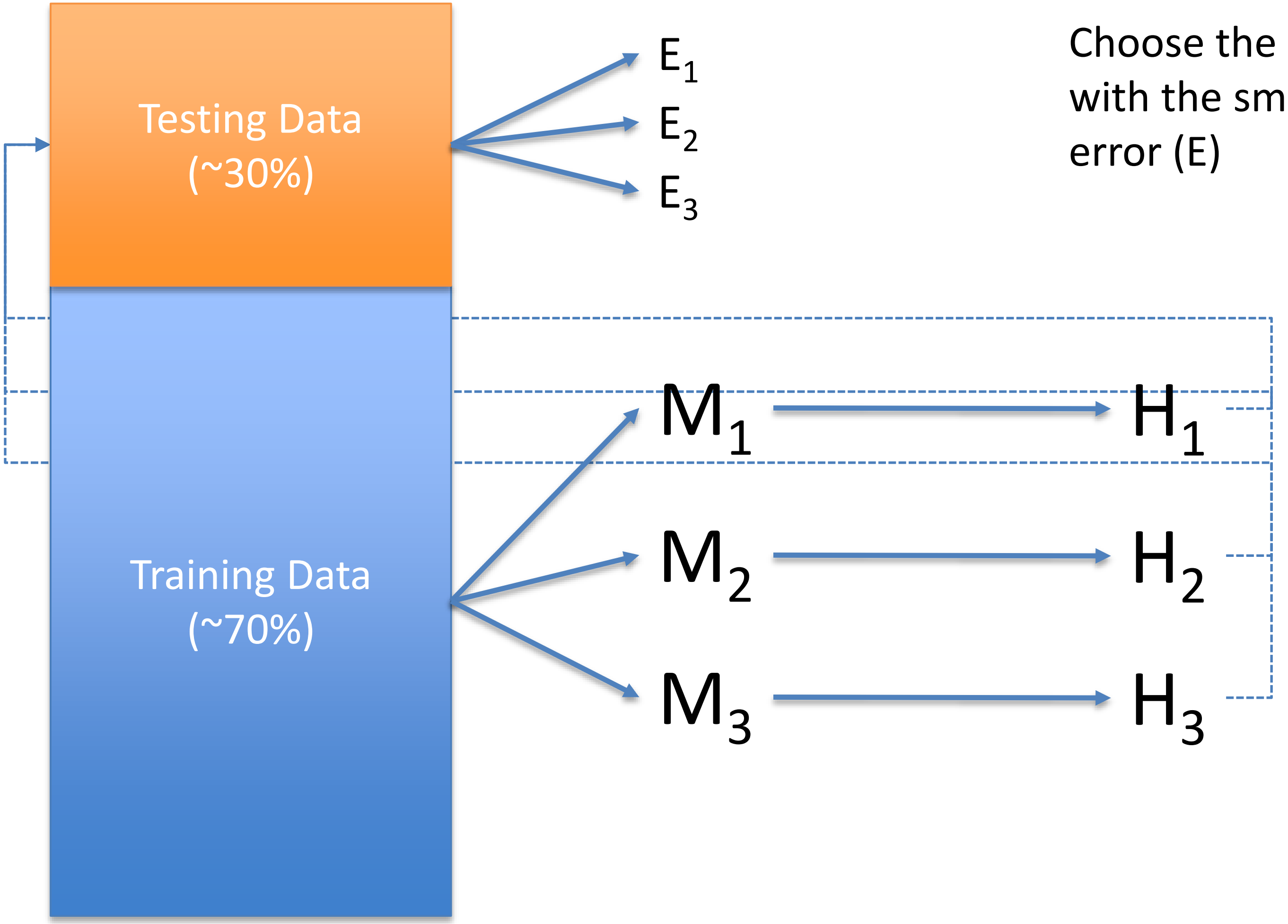
Validation Methods

- Cross validation
  - Test different models
  - Obtain reliable statistics
- Bias -- Variance Analysis
  - Regularization
  - Overfitting

Cross Validation

- The hypothesis with the smallest training error, won’t be the best.
  - Why?
  - We need test sets and training sets
- Our first tool is called hold-out cross validation.

Hold-out cross validation



Choose the one with the smallest error (E)

What is M

- Everything that we have assigned arbitrarily is fair game.
- Linear Regression
  - Order of the polynomial, regularization parameter
- SVM
  - Kernel, variables associated with kernel
- NN
  - Number of layers, activation functions, number of units.

Problems with Hold-out CV

- We are “wasting” ~70% of our data.

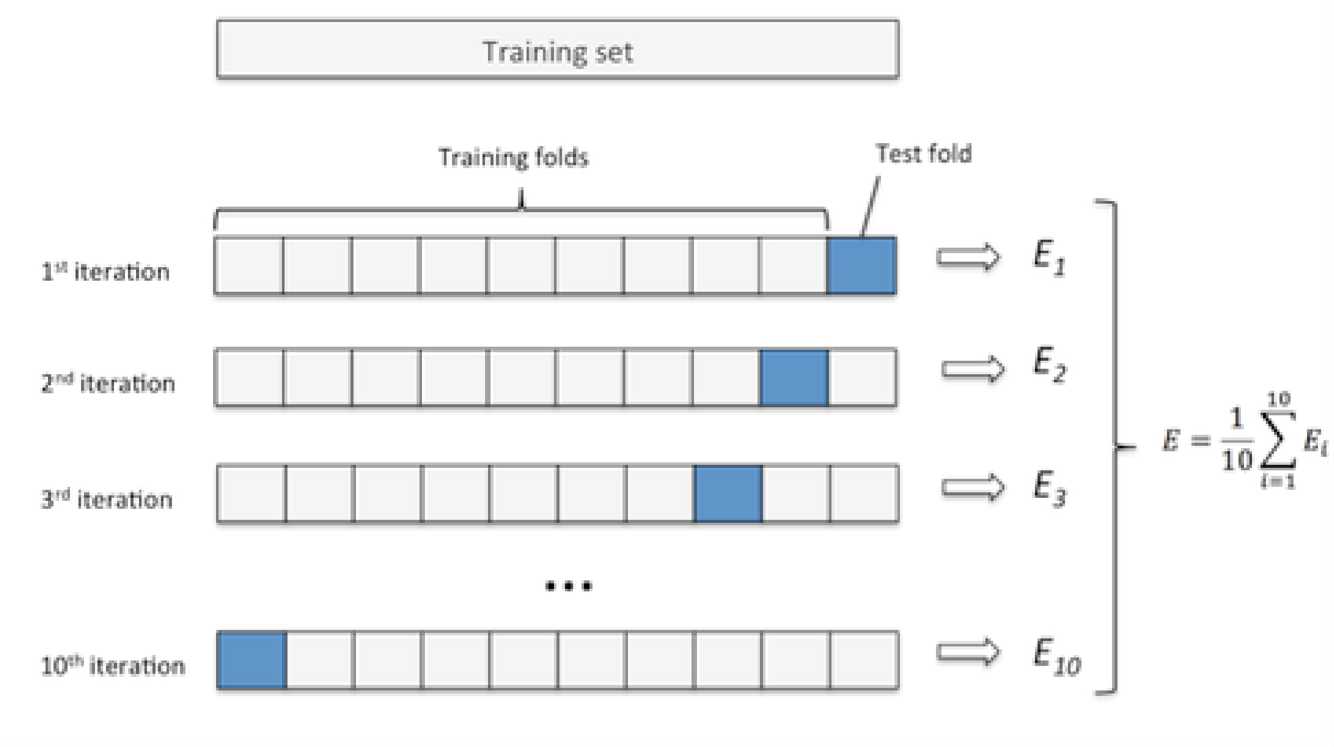
- For problems with few data points, this is just not desirable

- Be wary of papers that used CV, but have only few data points.
  - Be even more skeptic of papers that don’t mention CV at all.

An even better CV

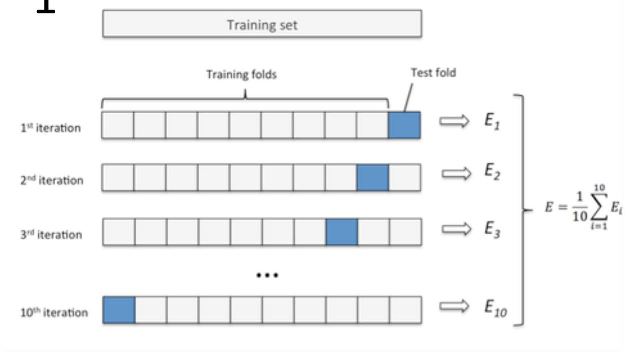
- K-fold CV
  - Split the data into k subsets (disjoint)
  - For each j = 1..k
    - Train model (M_i) in every subset, except j
    - Get an error (E_{ij}) for Model i in iteration j
  - Total error for M_i is going to be the average of all the errors (E_{ij})

K-Fold Cross validation



K-Fold Cross validation

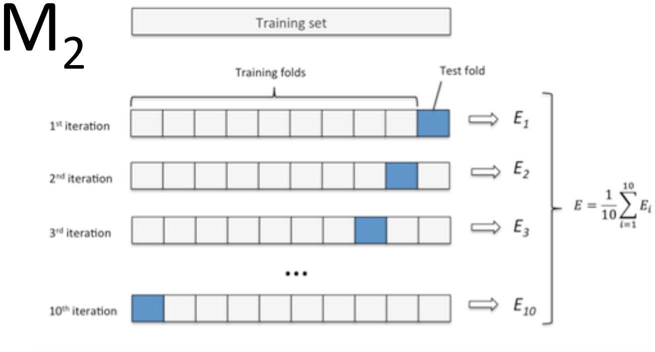
M_1



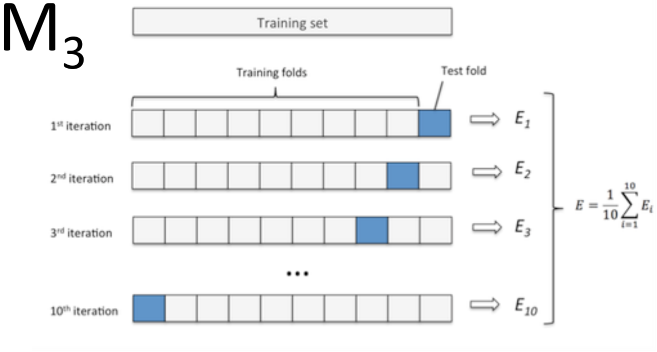
M_4

Training set
Training folds
Test fold
1st iteration
2nd iteration
3rd iteration
...
10th iteration
E_1
E_2
E_3
E_{10}
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$

M_2



M_3


Training set
Training folds
Test fold
1st iteration
2nd iteration
3rd iteration
...
10th iteration
E_1
E_2
E_3
E_{10}
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$