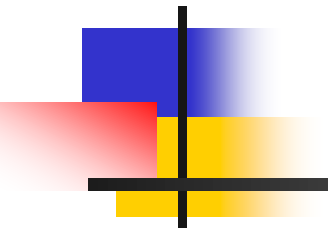


Model Essentials for Neural Net

Dr. Goutam Chakraborty

Outline

- Predict new cases.
- Select useful inputs.
- Optimize complexity.

Prediction formula

None

Stopped training

...

Model Essentials – Neural Networks

- Predict new cases.
- Select useful inputs.
- Optimize complexity.

Prediction formula
None
Stopped training

...

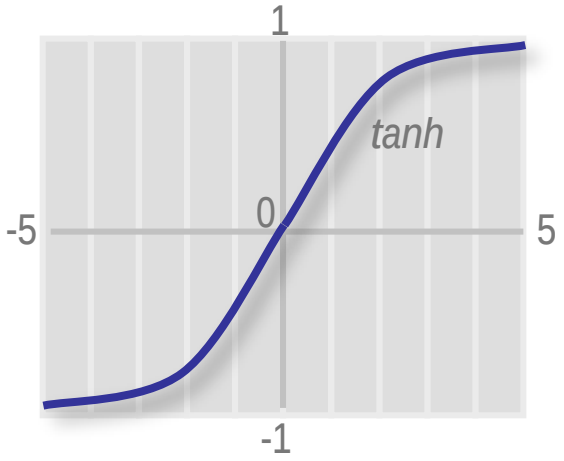
Neural Network Prediction Formula

hidden unit

prediction estimate

$$\hat{y} = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$

bias estimate *weight estimate*



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

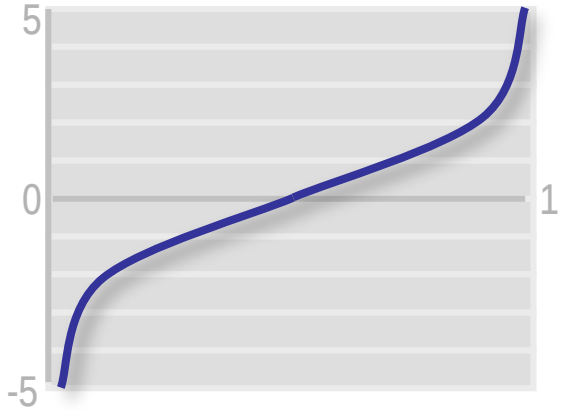
activation function

...

Neural Network Binary Prediction Formula

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

logit
link function



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

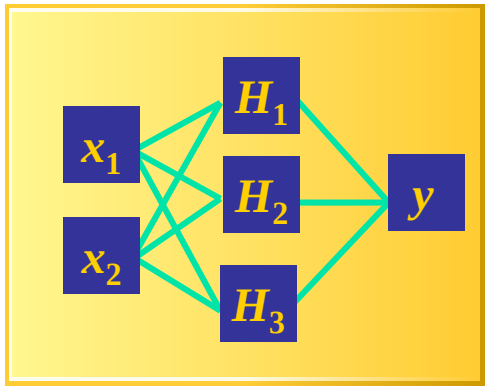
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

...

Neural Network Diagram

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$



input layer *hidden layer* *target layer*

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

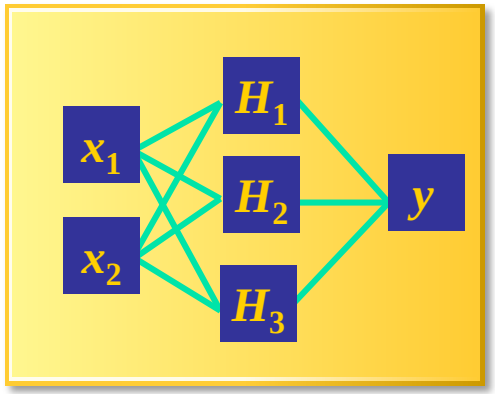
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

...

Neural Network Diagram

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$



$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

...

Prediction Illustration – Neural Networks

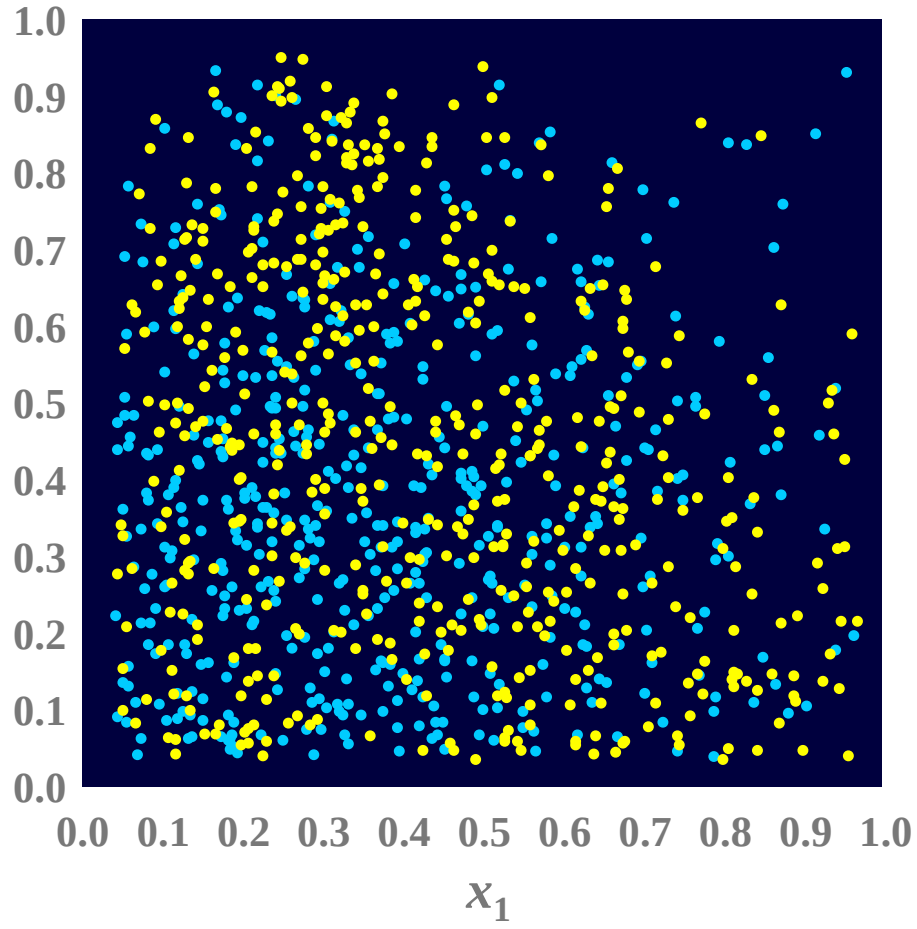
logit equation

$$\text{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$



...

Prediction Illustration – Neural Networks

logit equation

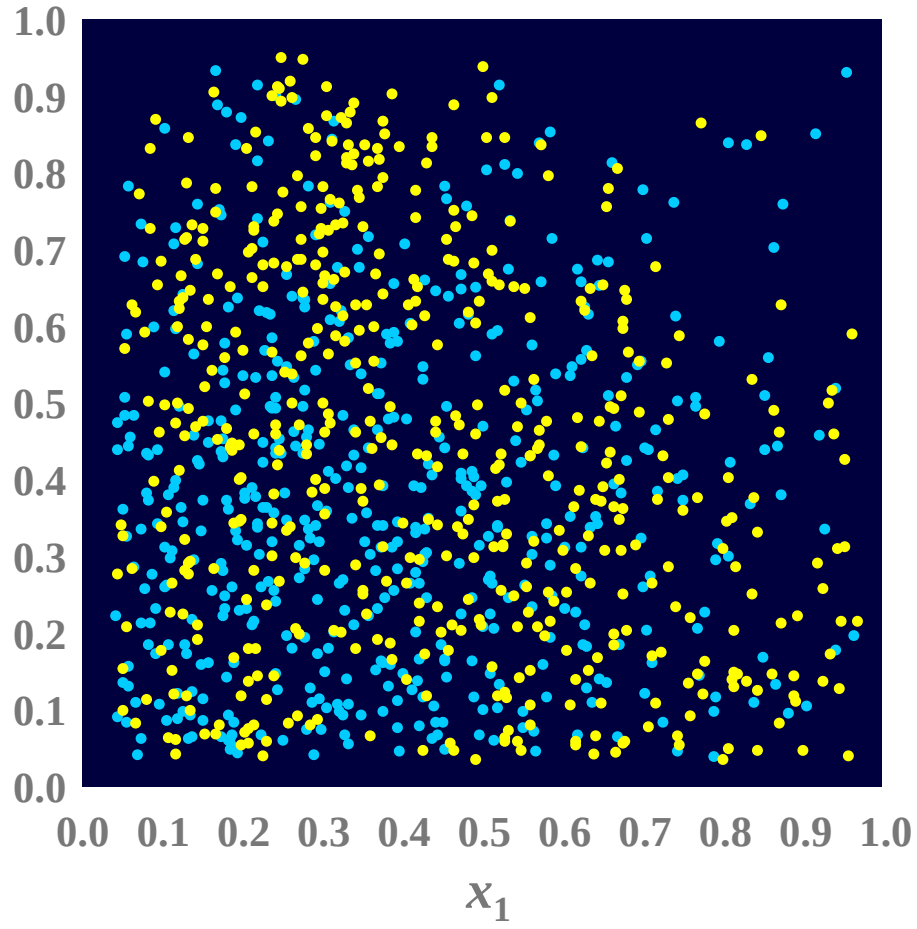
$$\text{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

Need weight estimates.



...

Prediction Illustration – Neural Networks

logit equation

$$\text{logit}(\hat{p}) = \hat{w}_{00} + \hat{w}_{01} H_1 + \hat{w}_{02} H_2 + \hat{w}_{03} H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

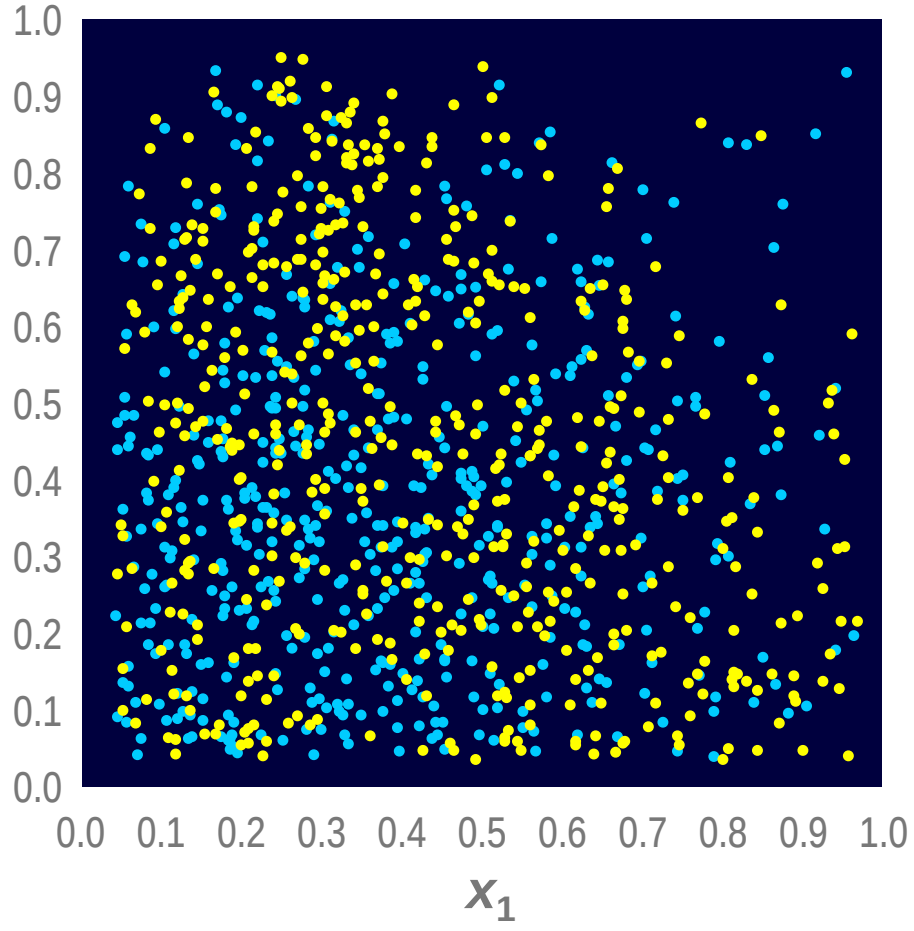
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$

Weight estimates are found by *maximizing*:

$$\sum_{\substack{\text{primary} \\ \text{outcome} \\ \text{training cases}}} \log(\hat{p}_i) + \sum_{\substack{\text{secondary} \\ \text{outcome} \\ \text{training cases}}} \log(1 - \hat{p}_i)$$

Log-likelihood Function



...

Prediction Illustration – Neural Networks

logit equation

$$\text{logit}(\hat{p}) = -0.5 + -2.6\ H_1 + -1.9\ H_2 + 0.63\ H_3$$

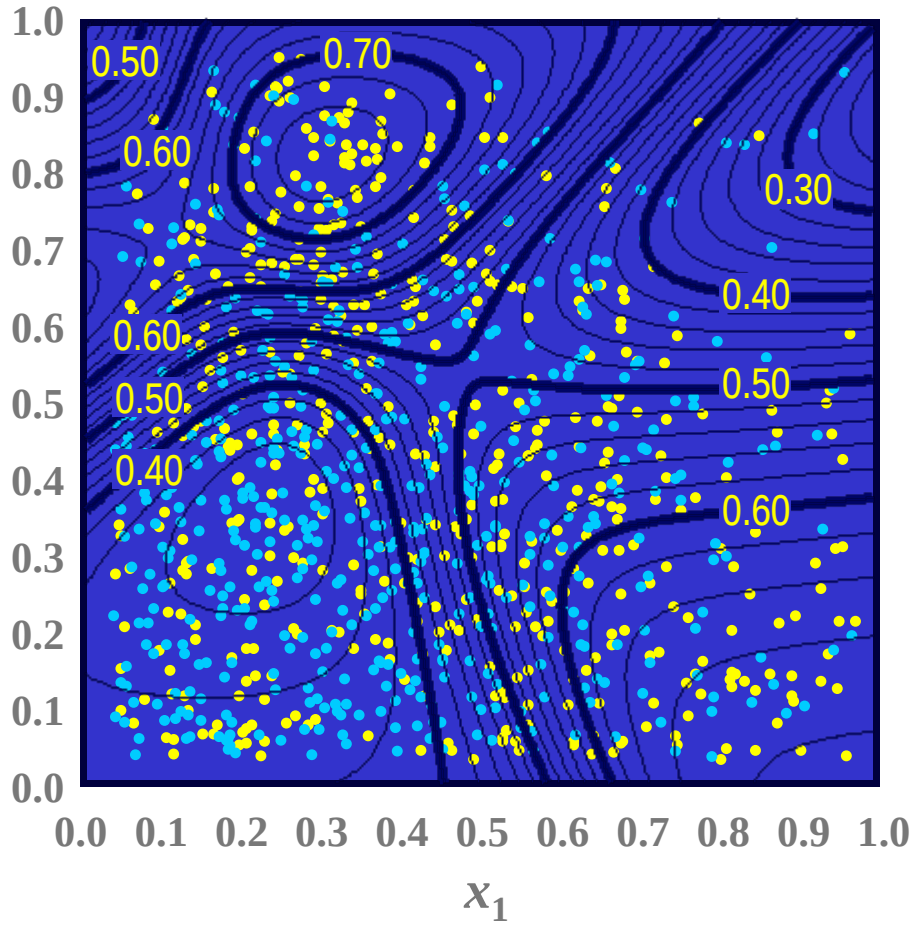
$$H_1 = \tanh(-1.8 + 0.25\ x_1 + -1.8\ x_2)$$

$$H_2 = \tanh(2.7 + 2.7\ x_1 + -5.3\ x_2)$$

$$H_3 = \tanh(-5.0 + 8.1\ x_1 + 4.3\ x_2)$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Probability estimates are obtained by solving the logit equation for $\hat{p}$ for each (x_1, x_2).



...

Neural Nets: Beyond the Prediction Formula

- Manage missing values.
- Handle extreme or unusual values.
- Use non-numeric inputs.
- Account for nonlinearities.
- Interpret the model.

...

Model Essentials – Neural Networks

- Predict new cases.
- Select useful inputs.
- Optimize complexity.

Prediction formula

None

Best model from sequence

Model Essentials – Neural Networks

- Predict new cases.
- Select useful inputs.
- Optimize complexity.

Prediction formula

Sequential selection

Stopped training

...

Fit Statistic versus Optimization Iteration

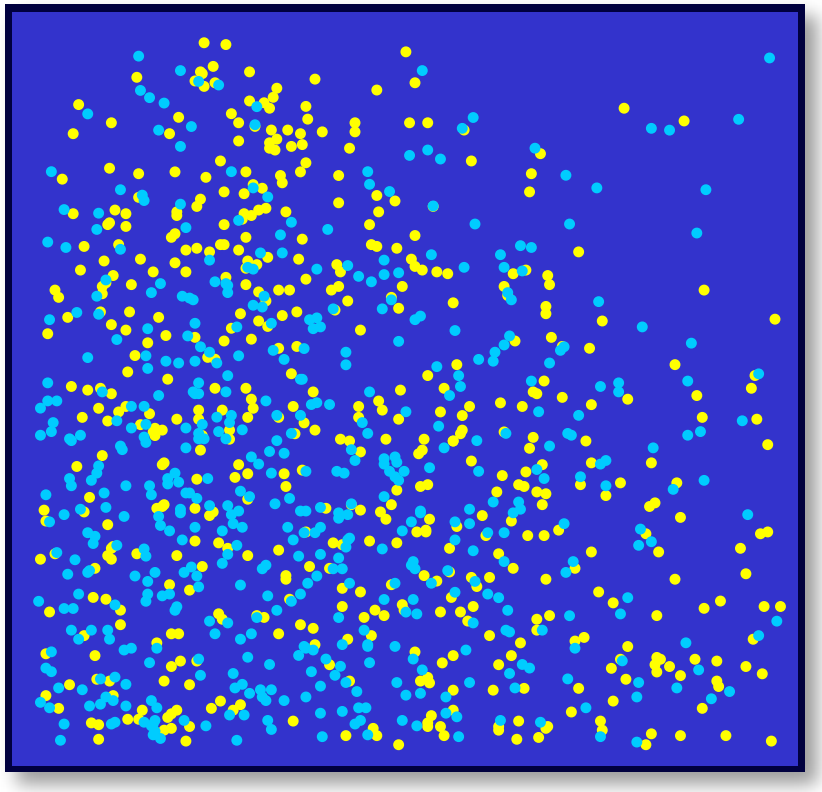
initial hidden unit weights

$$\text{logit}(\hat{p}) = 0 + 0 \cdot H_1 + 0 \cdot H_2 + 0 \cdot H_3$$

$$H_1 = \tanh(-1.5 - .03x_1 - .07x_2)$$

$$H_2 = \tanh(.79 - .17x_1 - .16x_2)$$

$$H_3 = \tanh(.57 + .05x_1 + .35x_2)$$



...

Fit Statistic versus Optimization Iteration

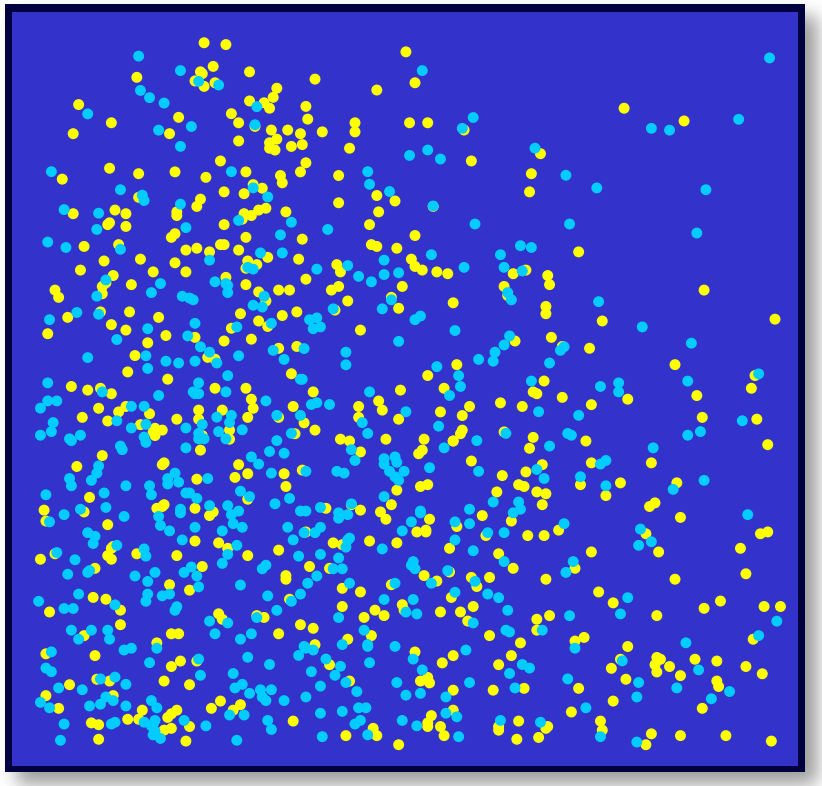
$$\text{logit}(\hat{p}) = 0 + 0 \cdot H_1 + 0 \cdot H_2 + 0 \cdot H_3$$

$$H_1 = \tanh(-1.5 - .03x_1 - .07x_2)$$

$$H_2 = \tanh(.79 - .17x_1 - .16x_2)$$

$$H_3 = \tanh(.57 + .05x_1 + .35x_2)$$

random initial
input weights and biases



...

Fit Statistic versus Optimization Iteration

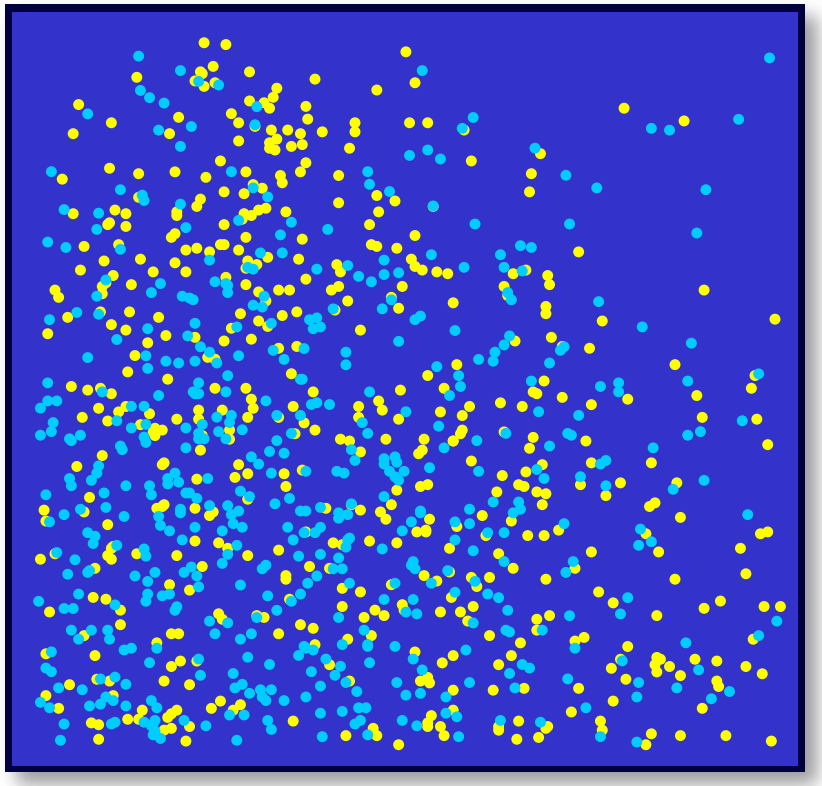
$$\text{logit}(\hat{p}) = 0 + 0 \cdot H_1 + 0 \cdot H_2 + 0 \cdot H_3$$

$$H_1 = \tanh(-1.5 - .03x_1 - .07x_2)$$

$$H_2 = \tanh(.79 - .17x_1 - .16x_2)$$

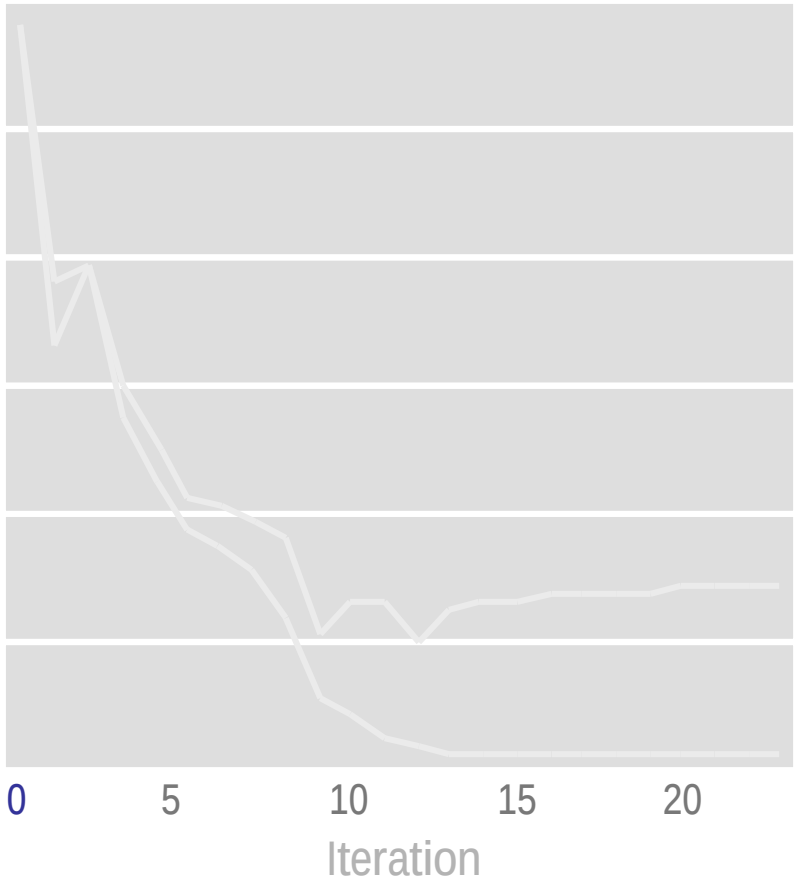
$$H_3 = \tanh(.57 + .05x_1 + .35x_2)$$

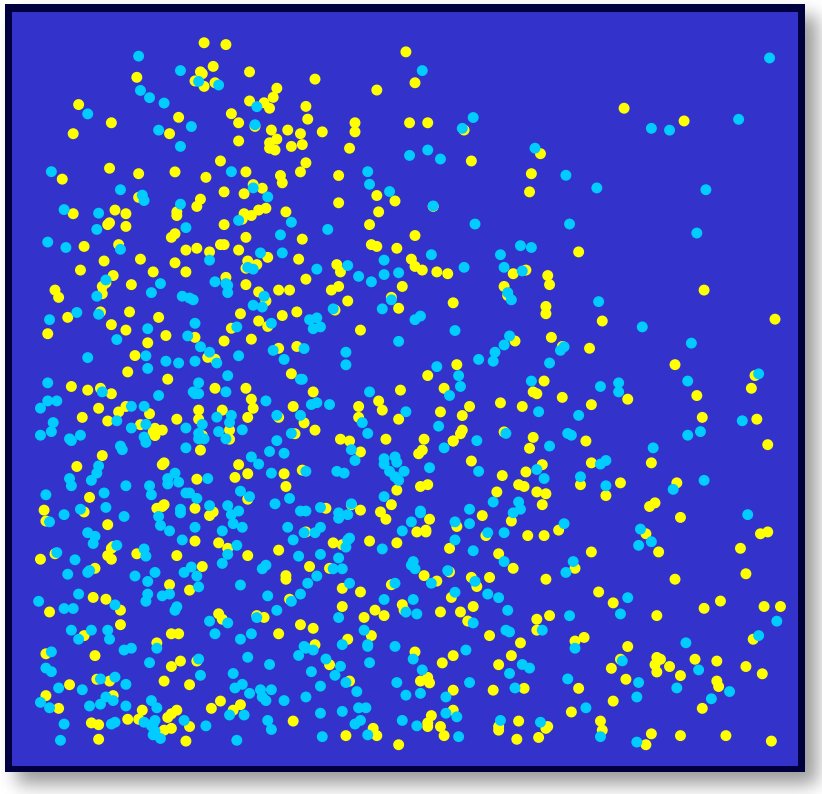
random initial
input weights and biases



...

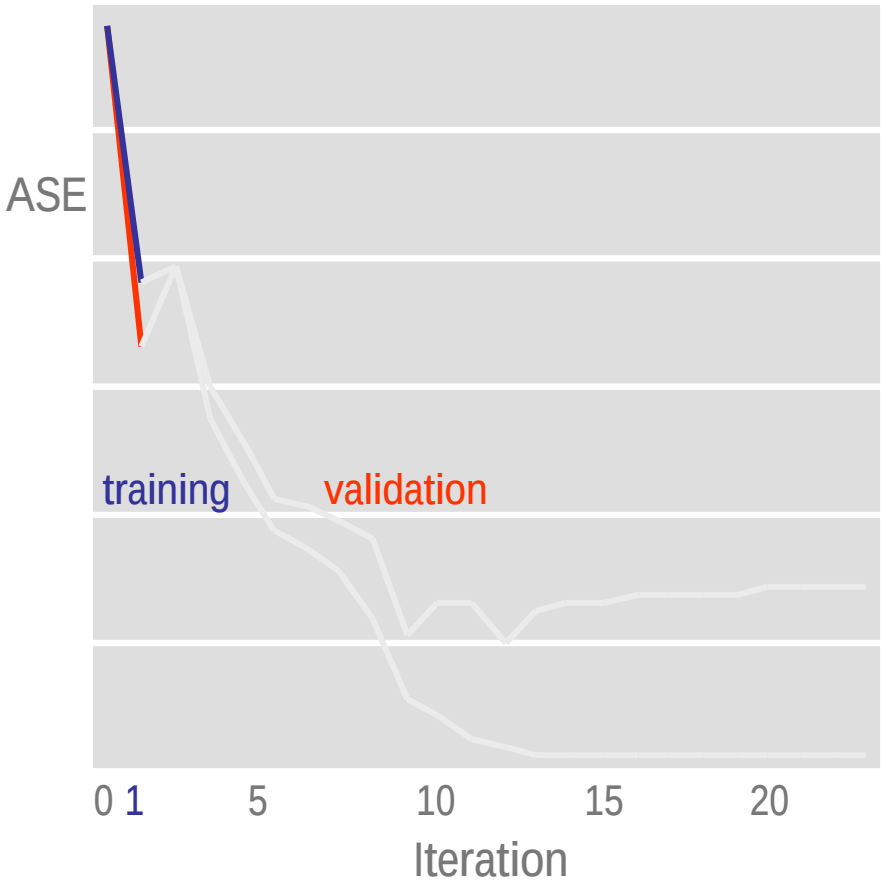
Fit Statistic versus Optimization Iteration

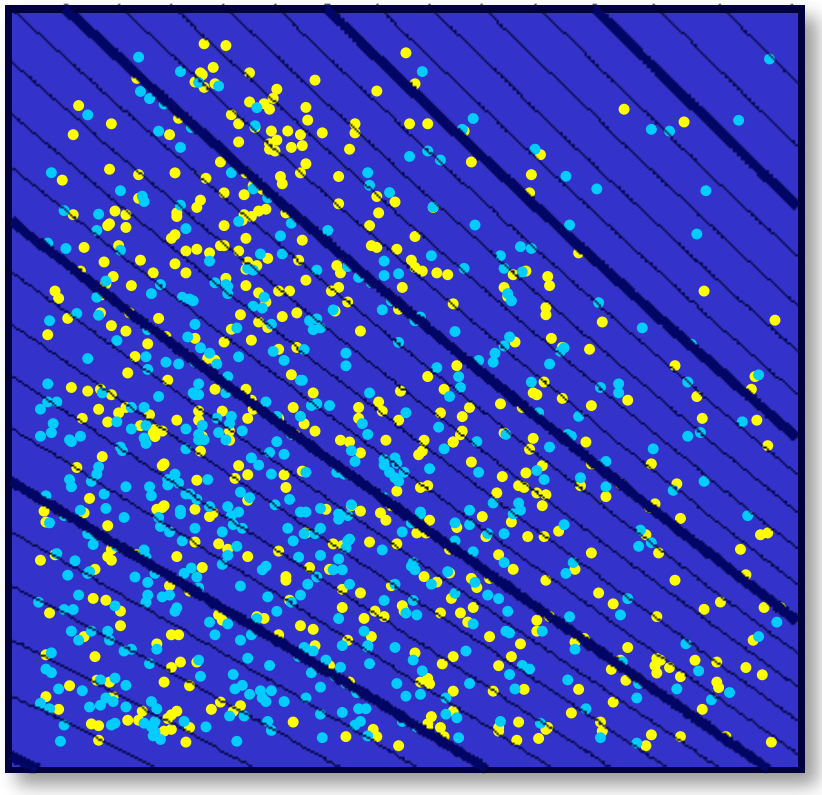




...

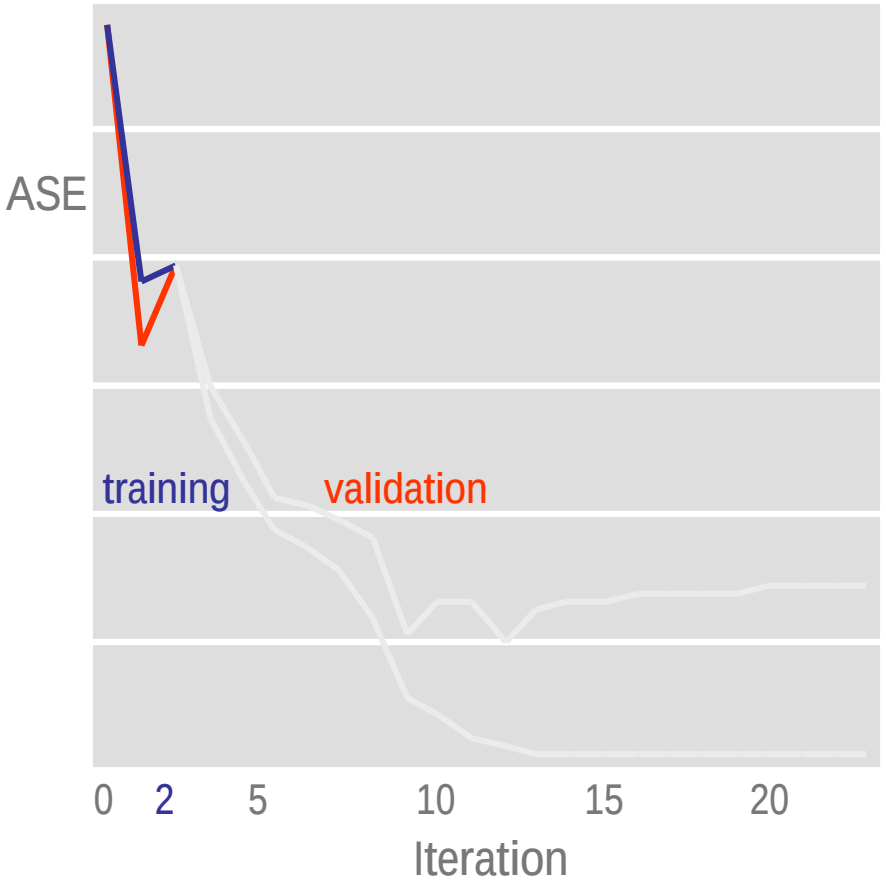
Fit Statistic versus Optimization Iteration

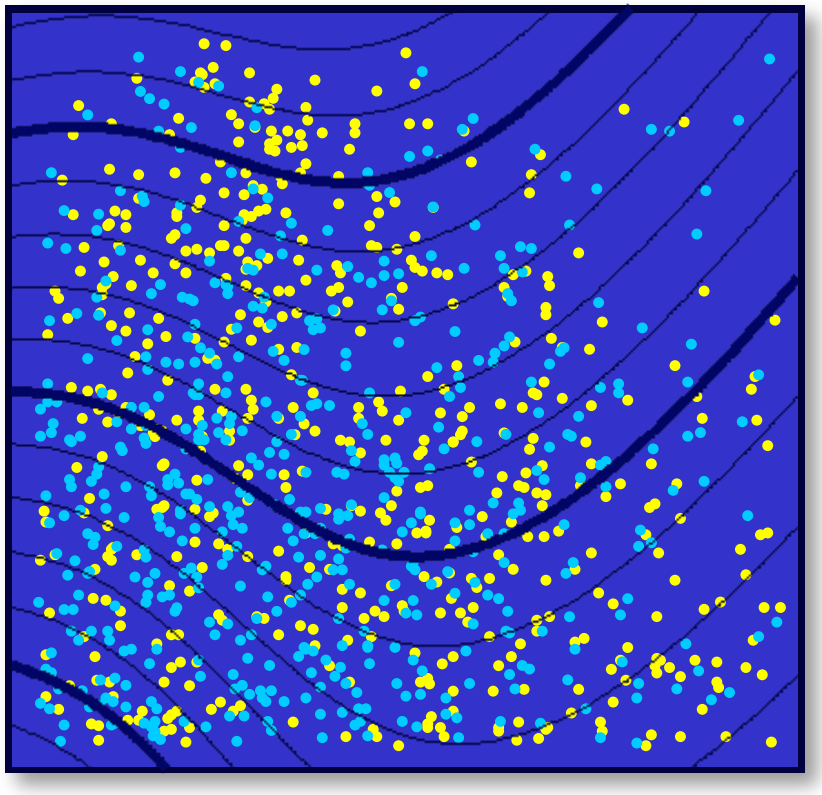




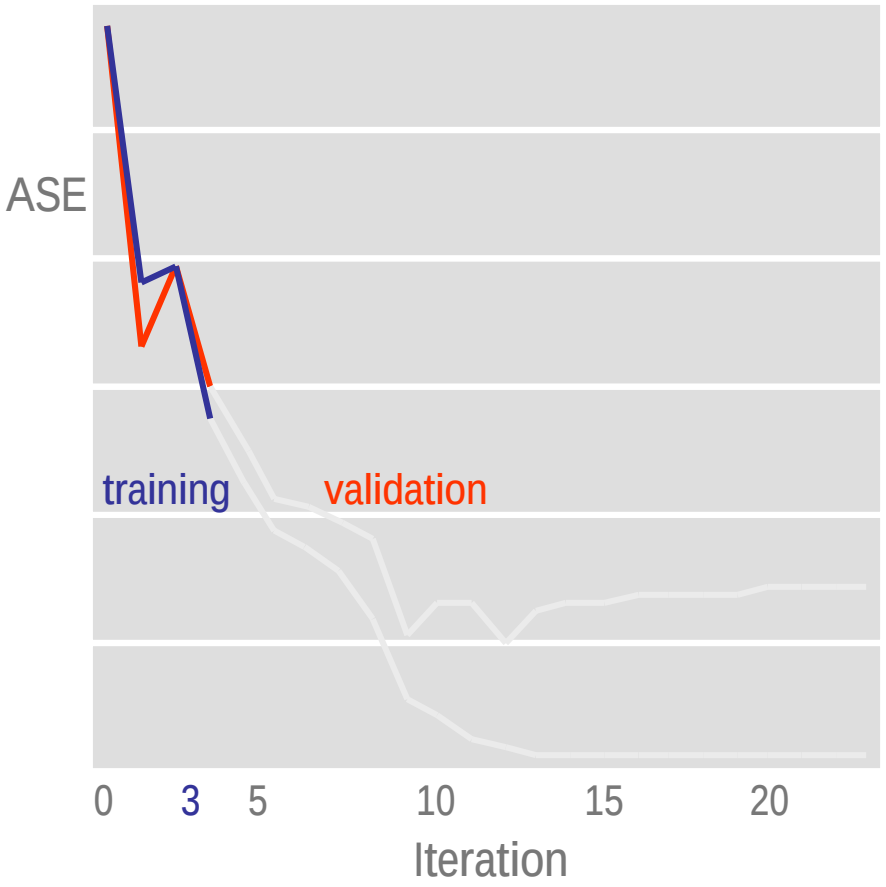
...

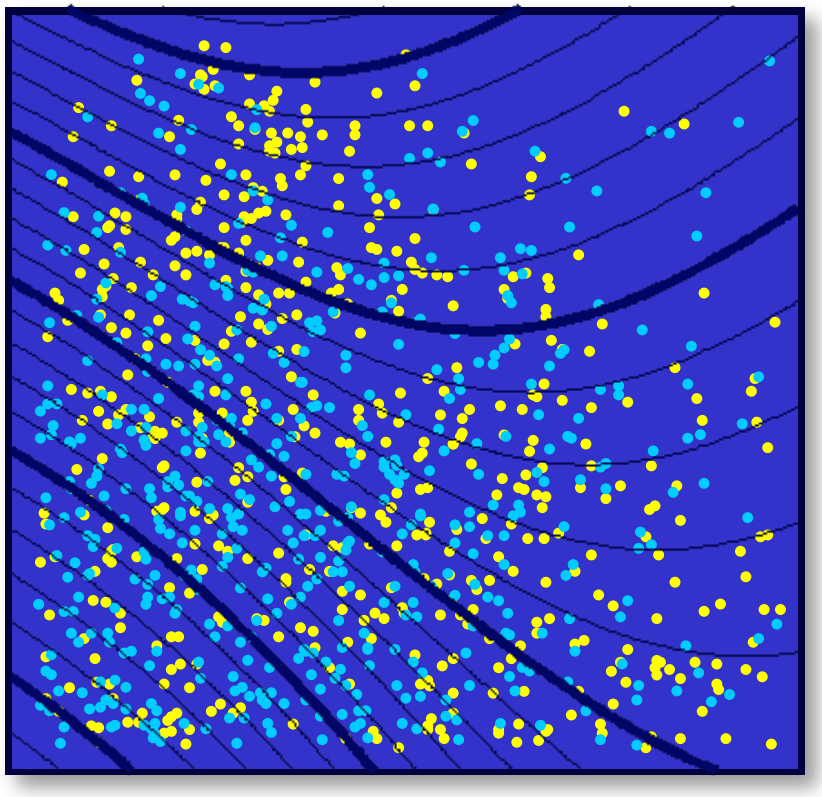
Fit Statistic versus Optimization Iteration



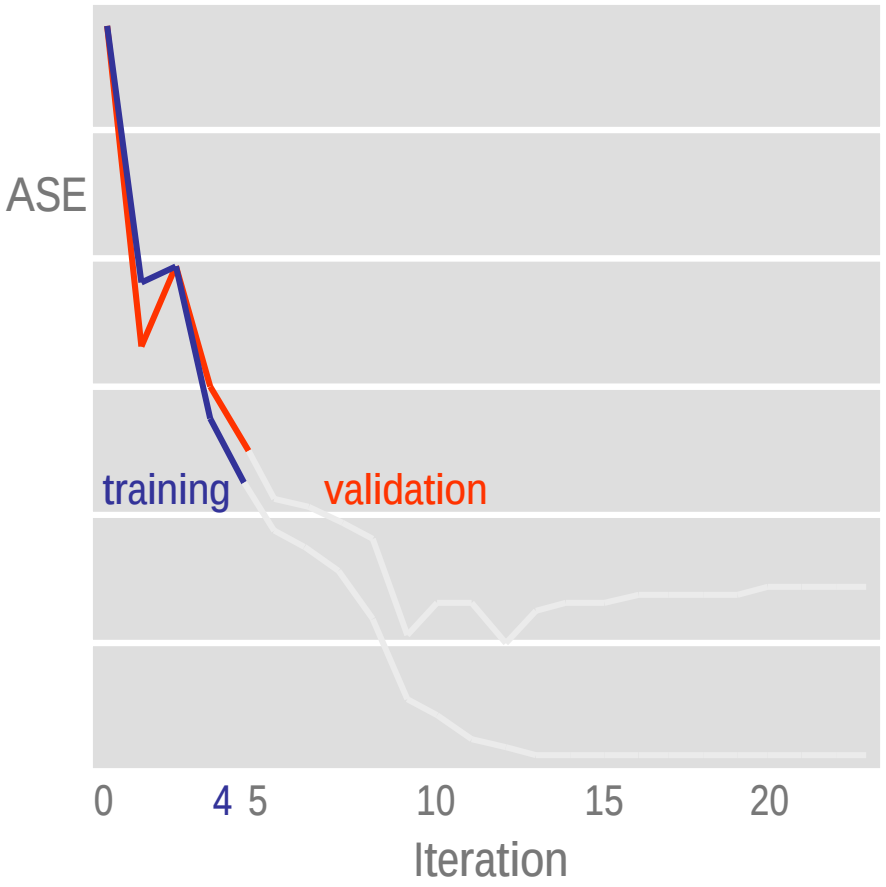


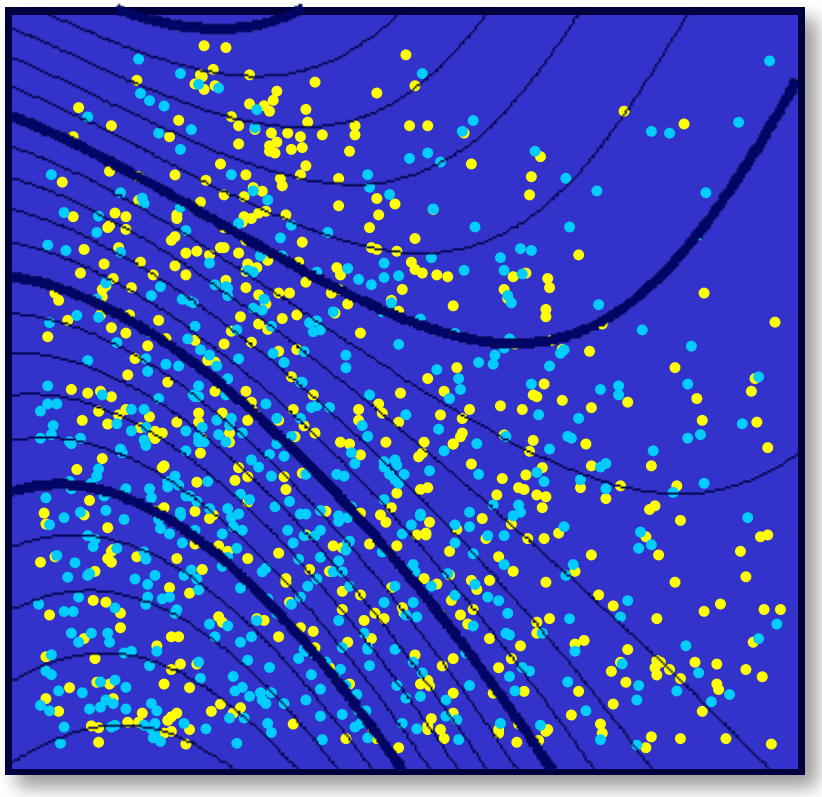
Fit Statistic versus Optimization Iteration



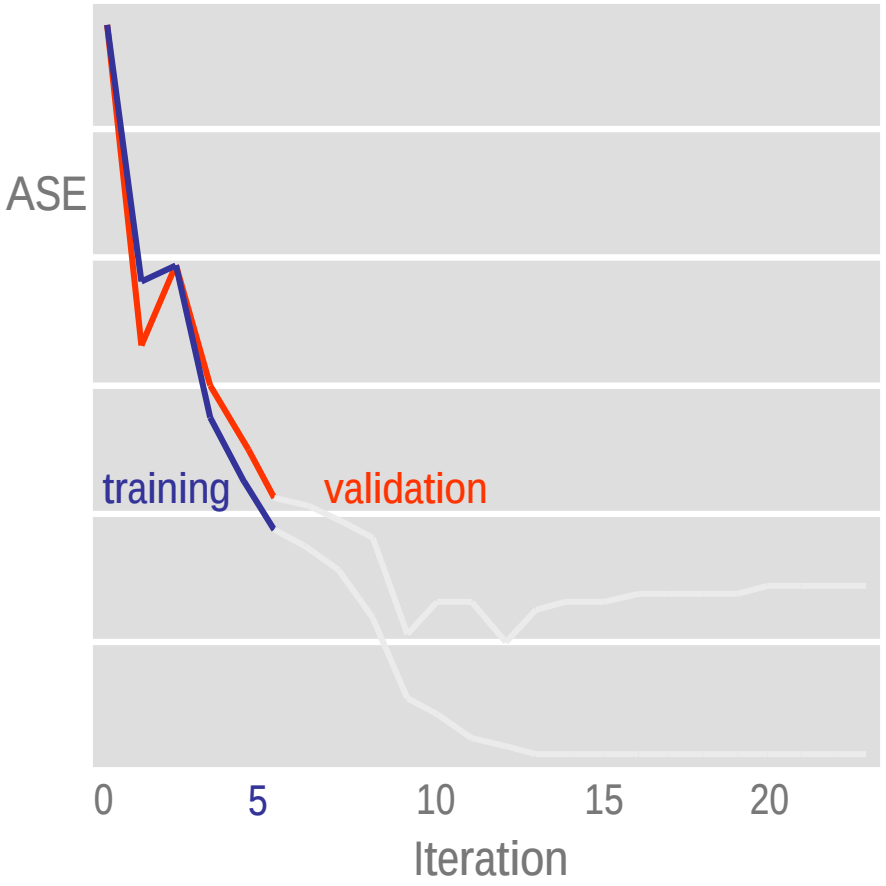


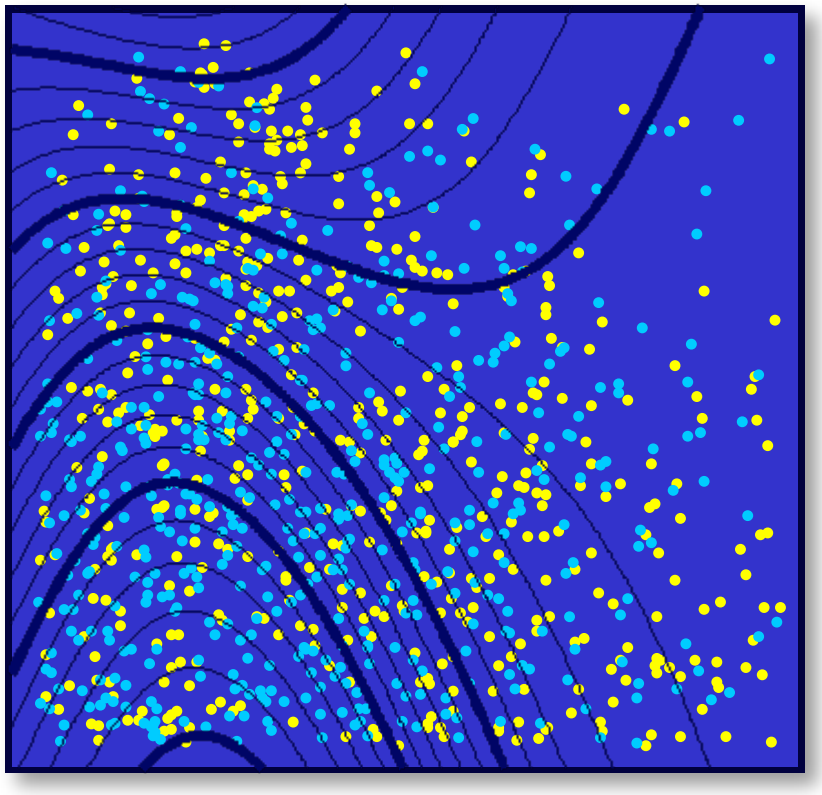
Fit Statistic versus Optimization Iteration



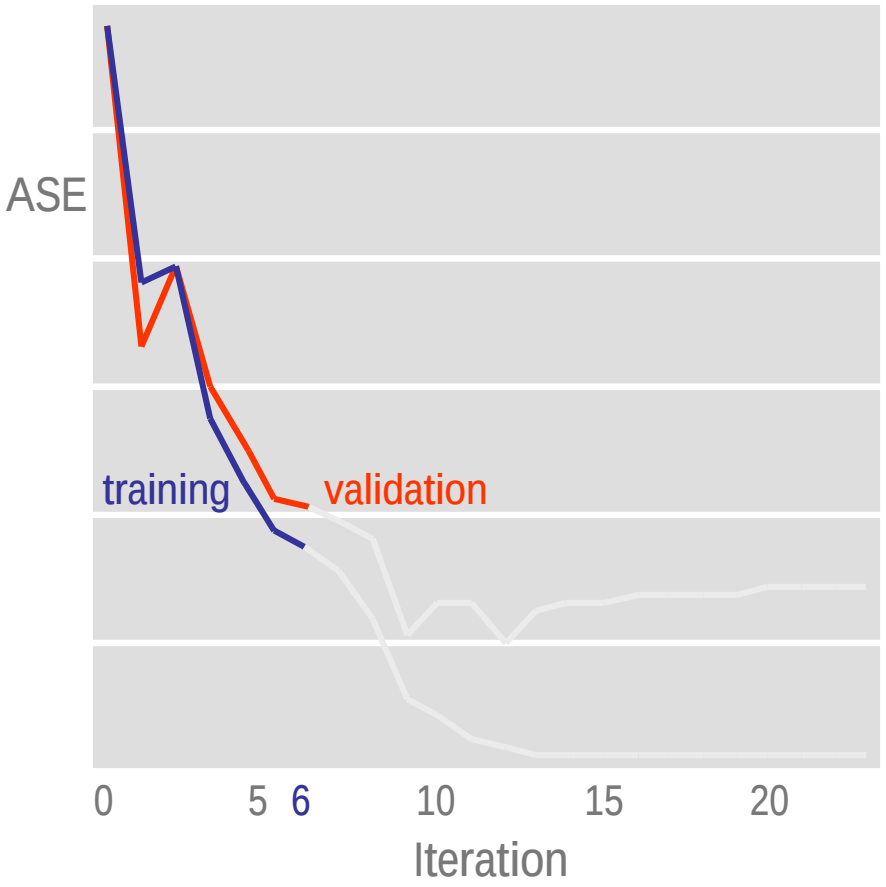


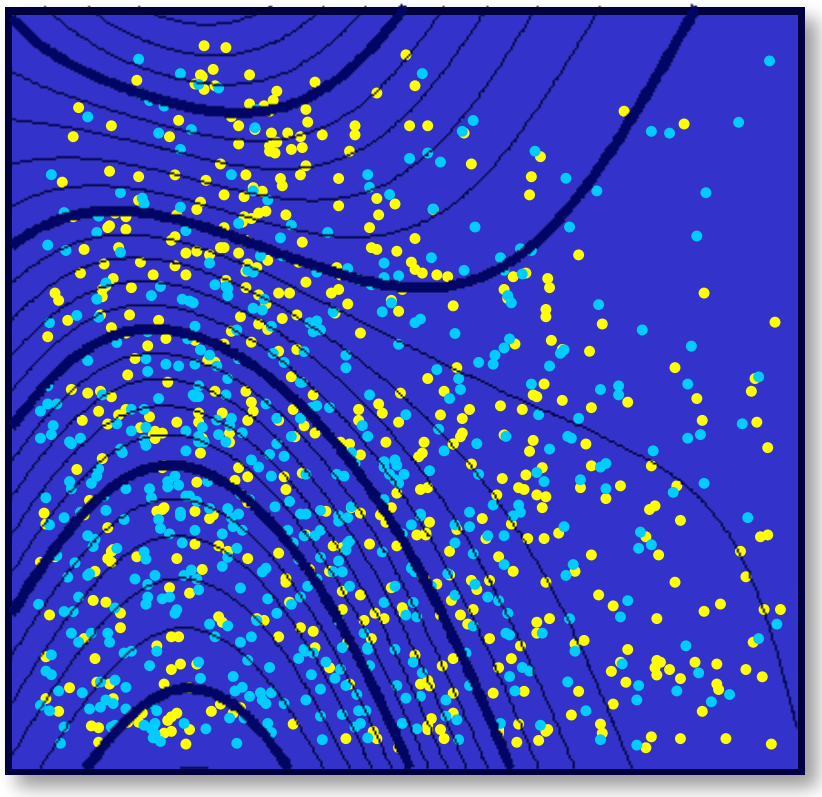
Fit Statistic versus Optimization Iteration





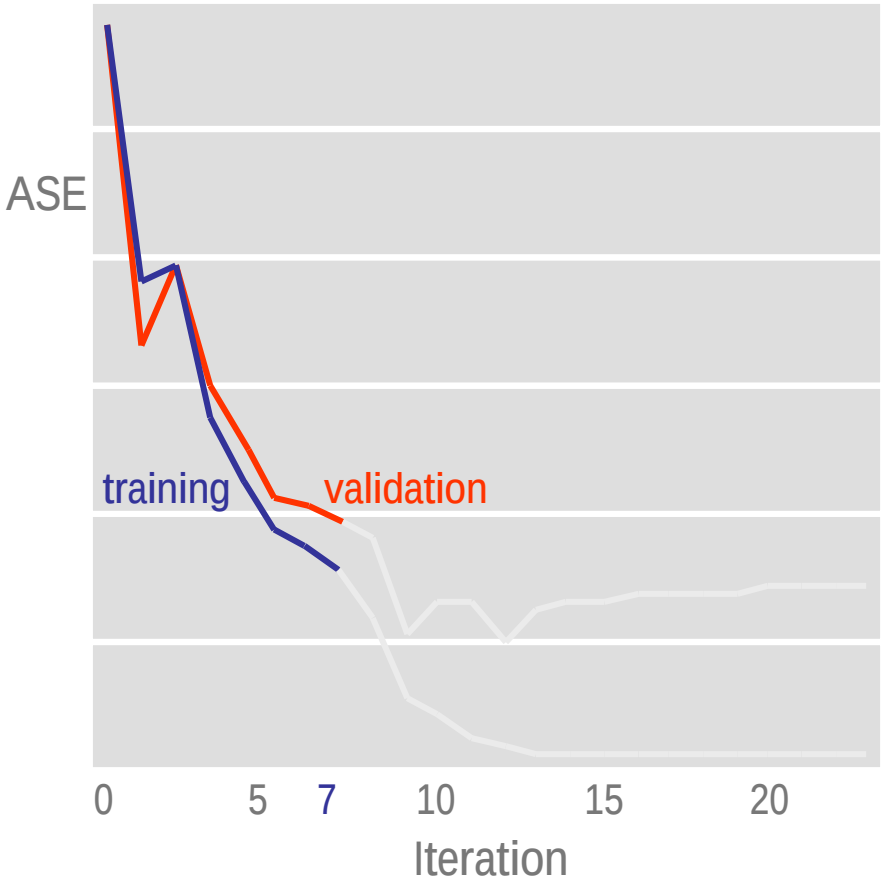
Fit Statistic versus Optimization Iteration

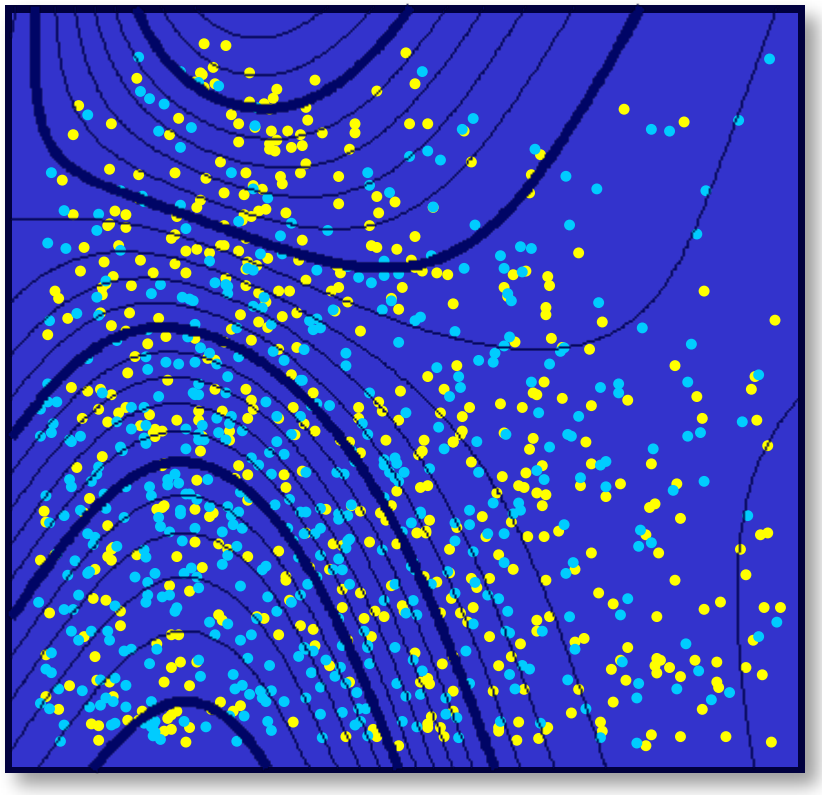




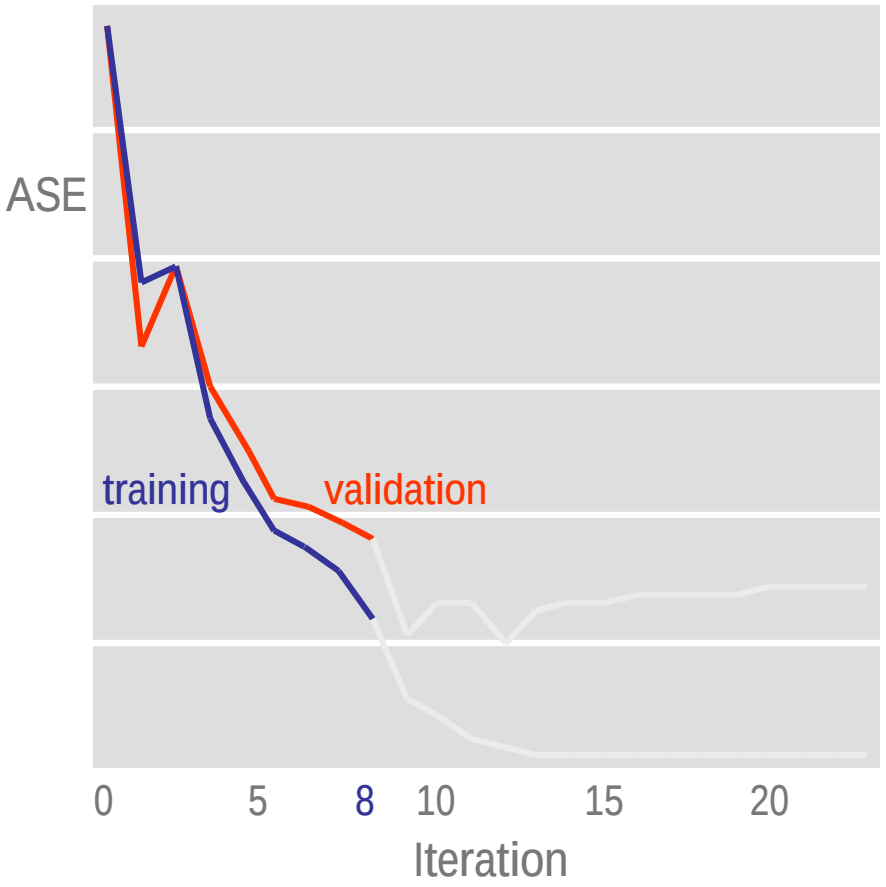
...

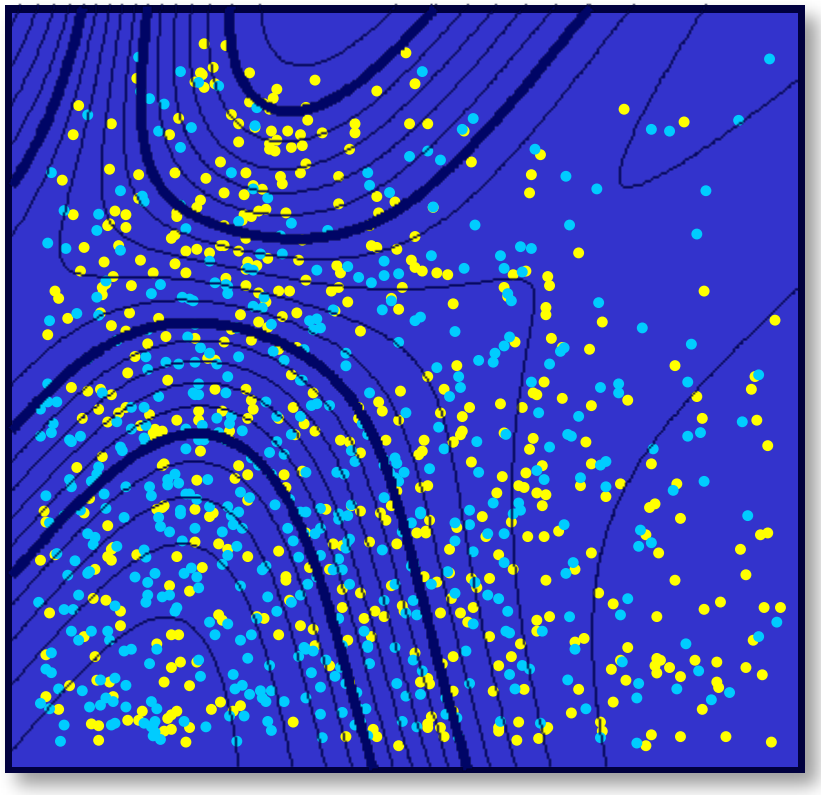
Fit Statistic versus Optimization Iteration





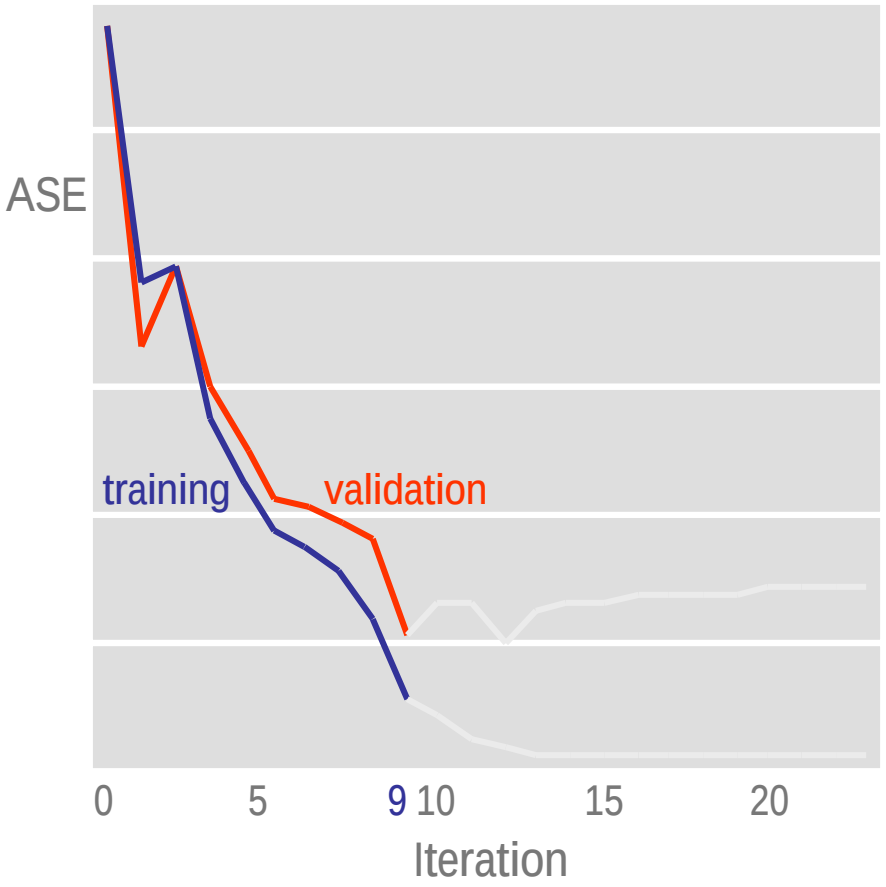
Fit Statistic versus Optimization Iteration

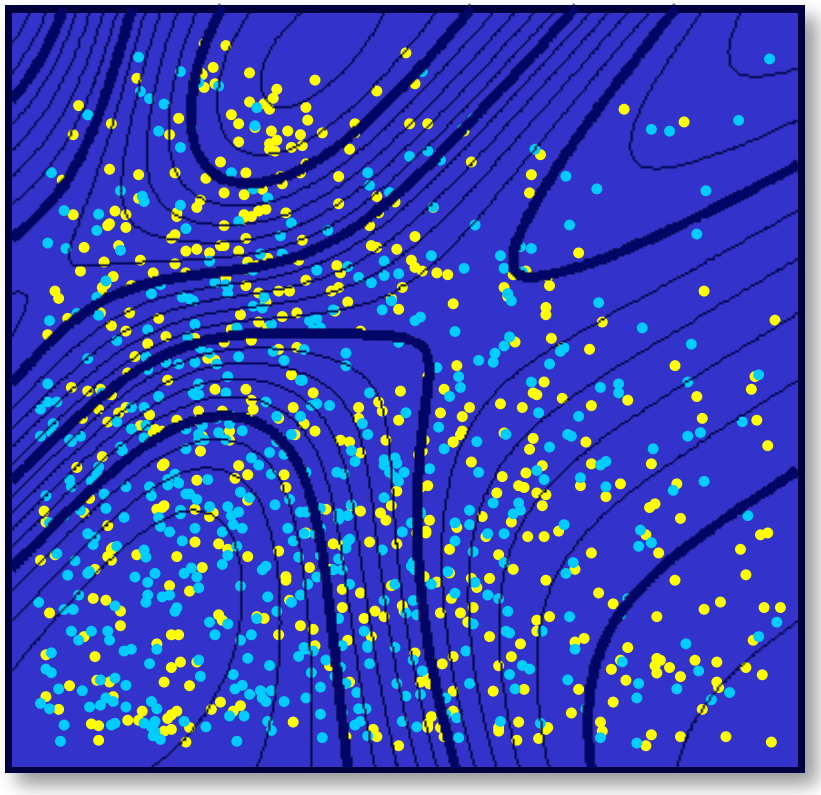




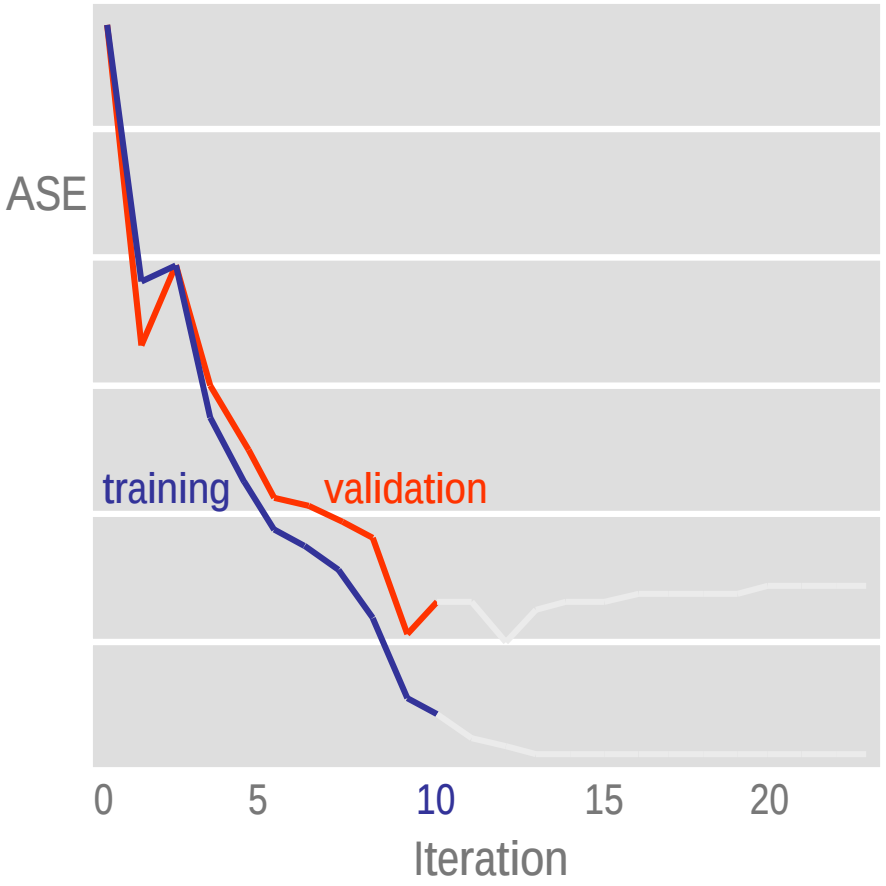
...

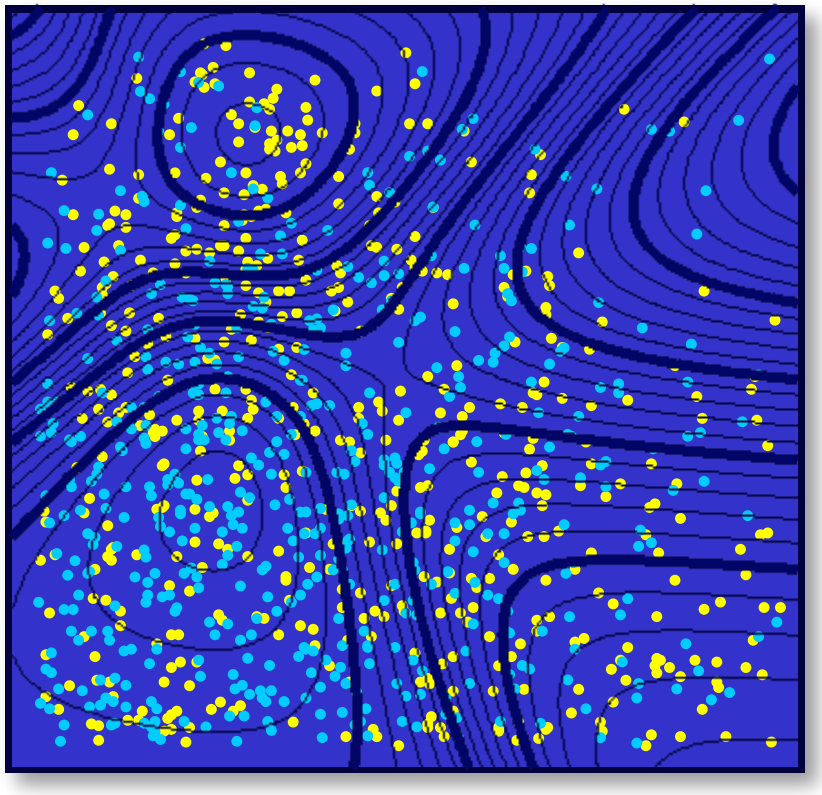
Fit Statistic versus Optimization Iteration



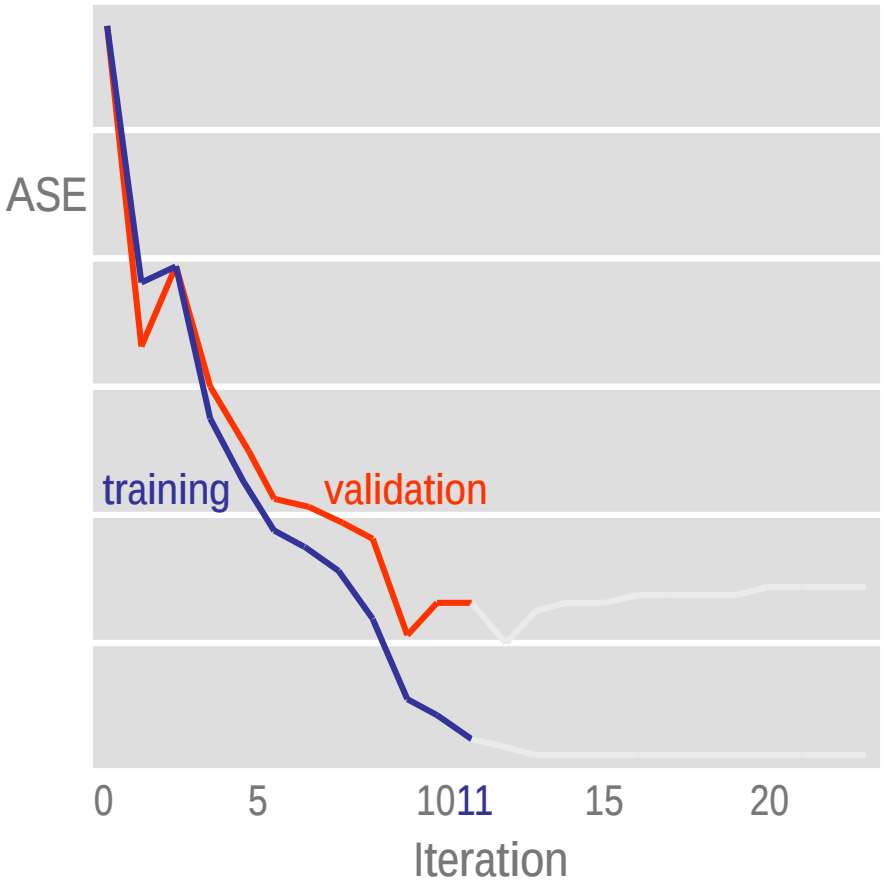


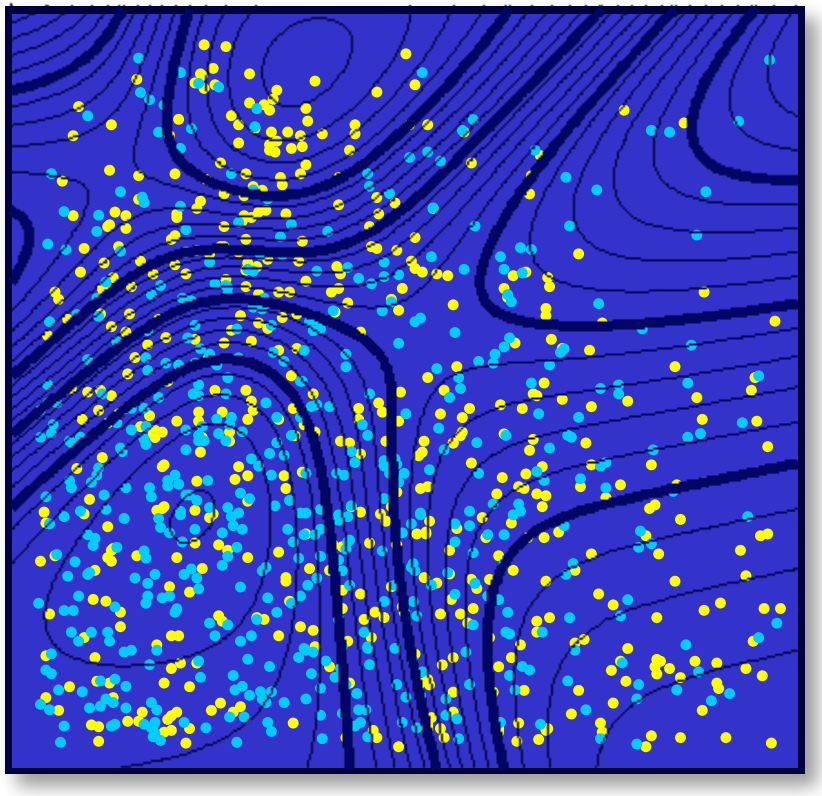
Fit Statistic versus Optimization Iteration





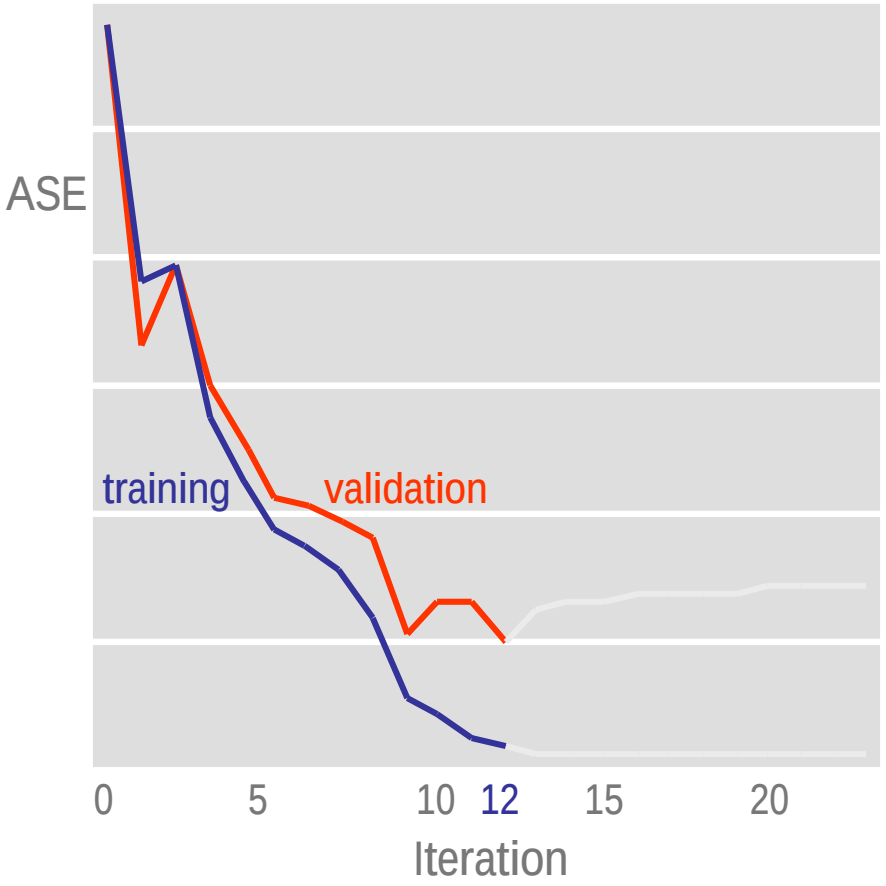
Fit Statistic versus Optimization Iteration

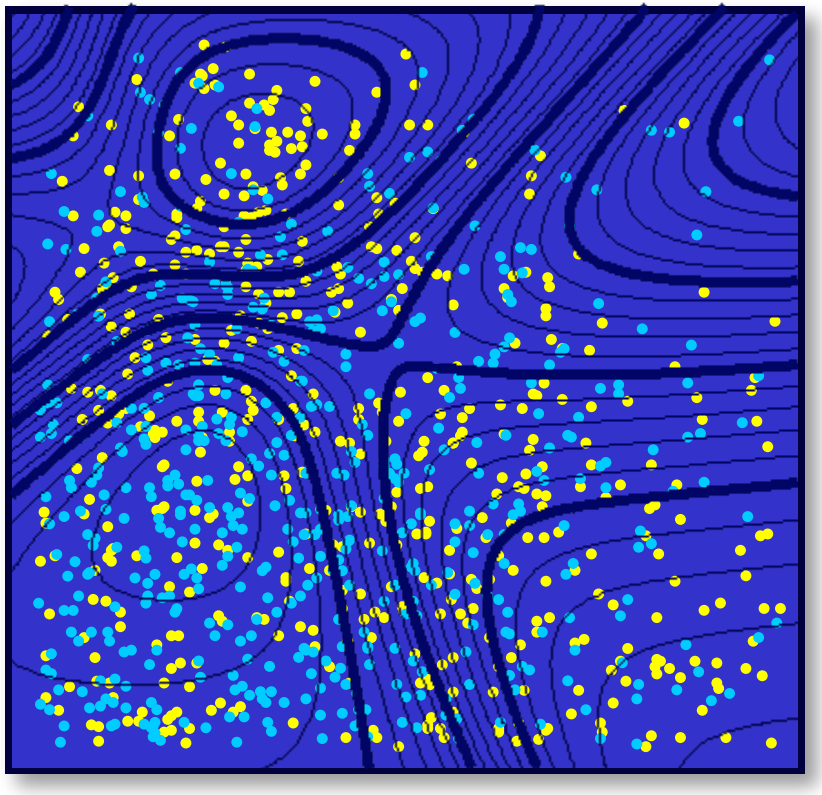




...

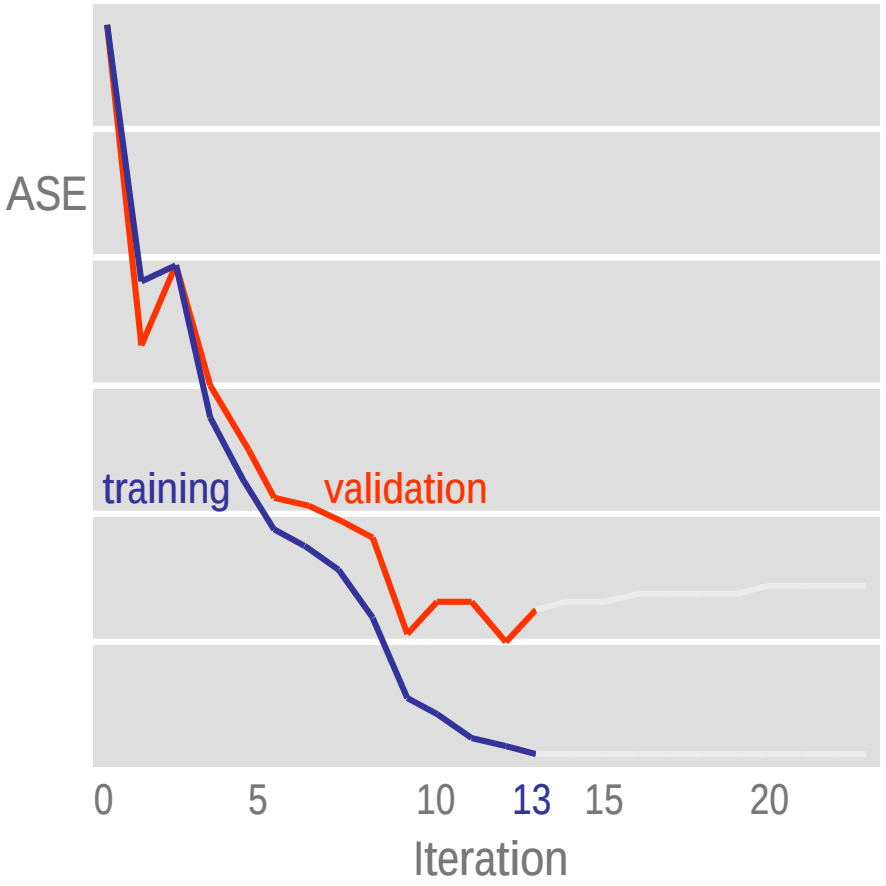
Fit Statistic versus Optimization Iteration

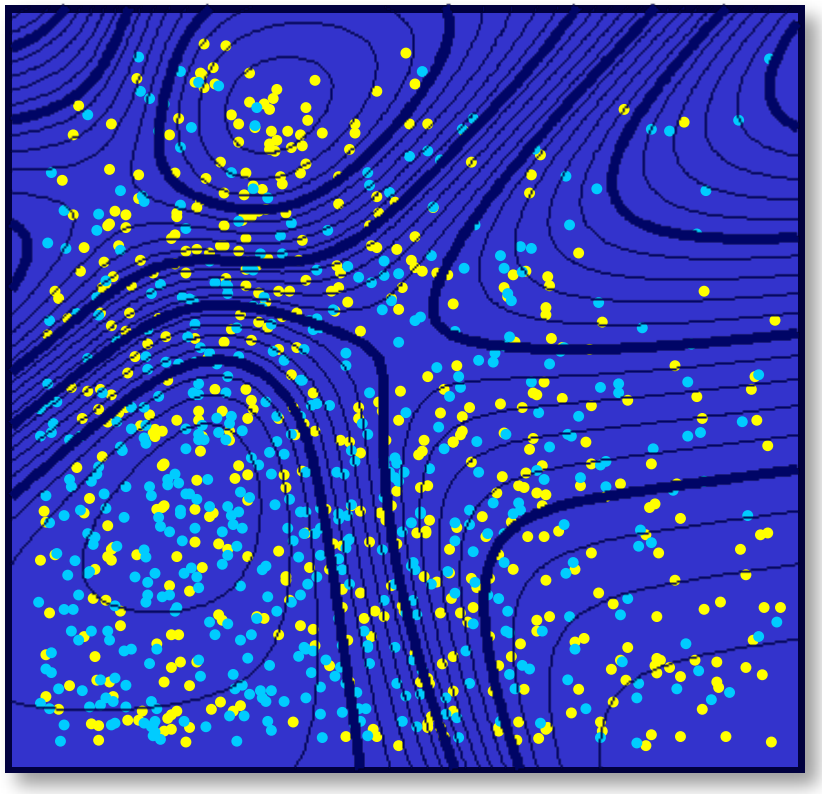




...

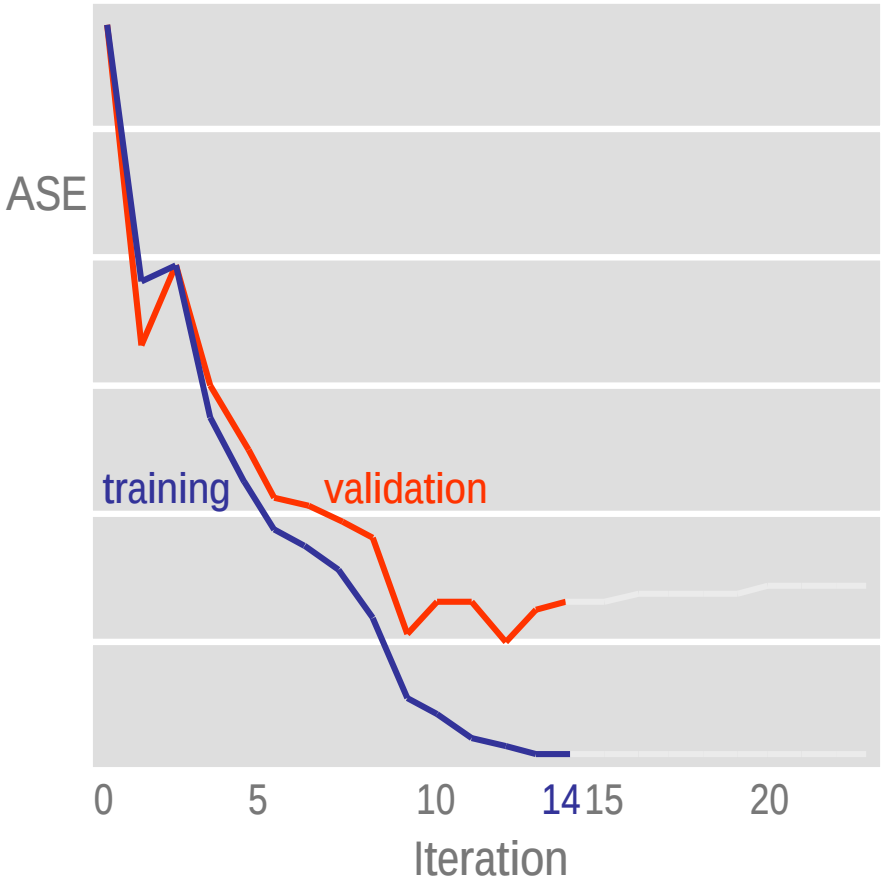
Fit Statistic versus Optimization Iteration

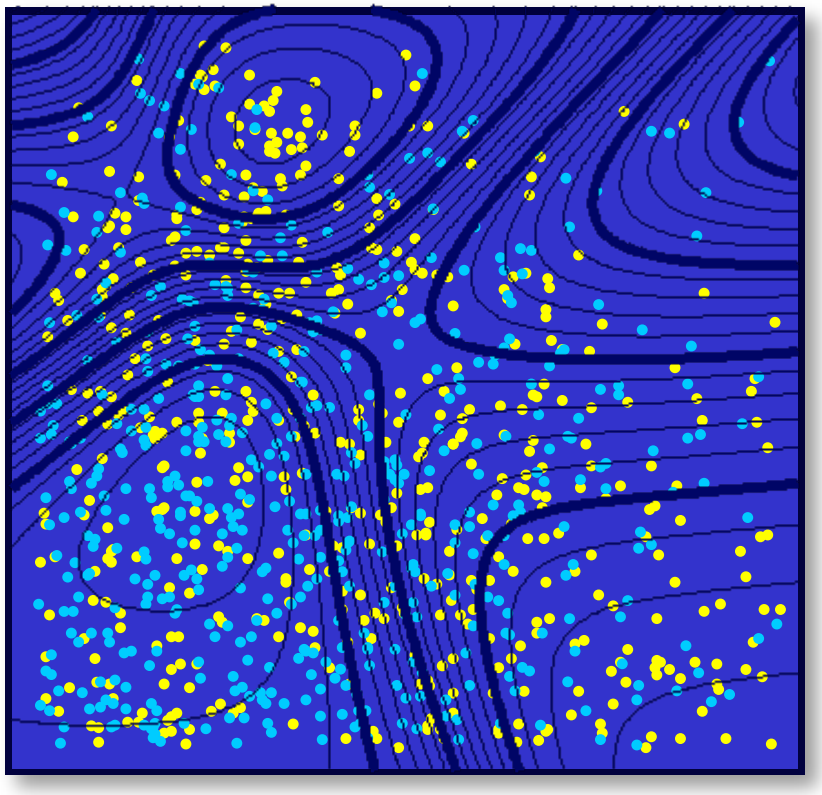




...

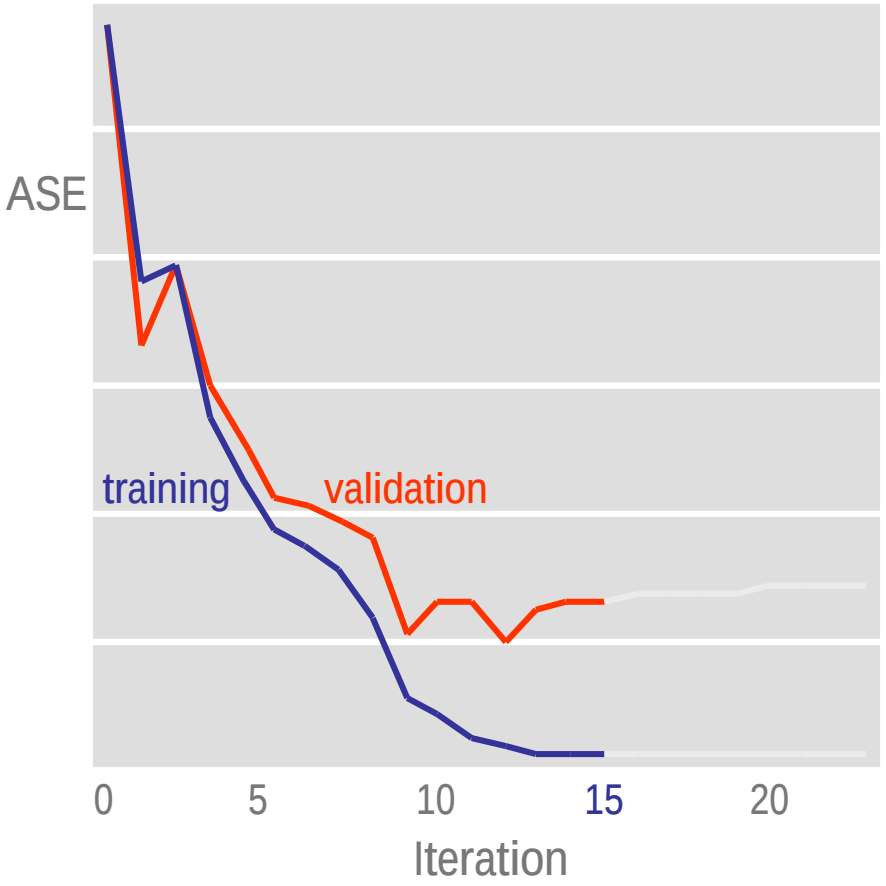
Fit Statistic versus Optimization Iteration

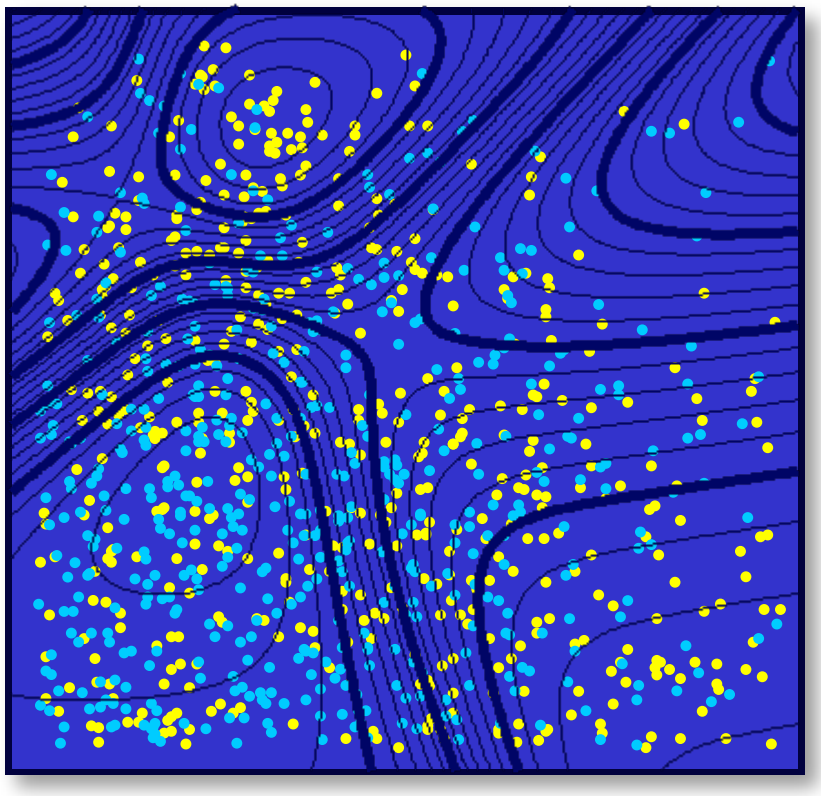




...

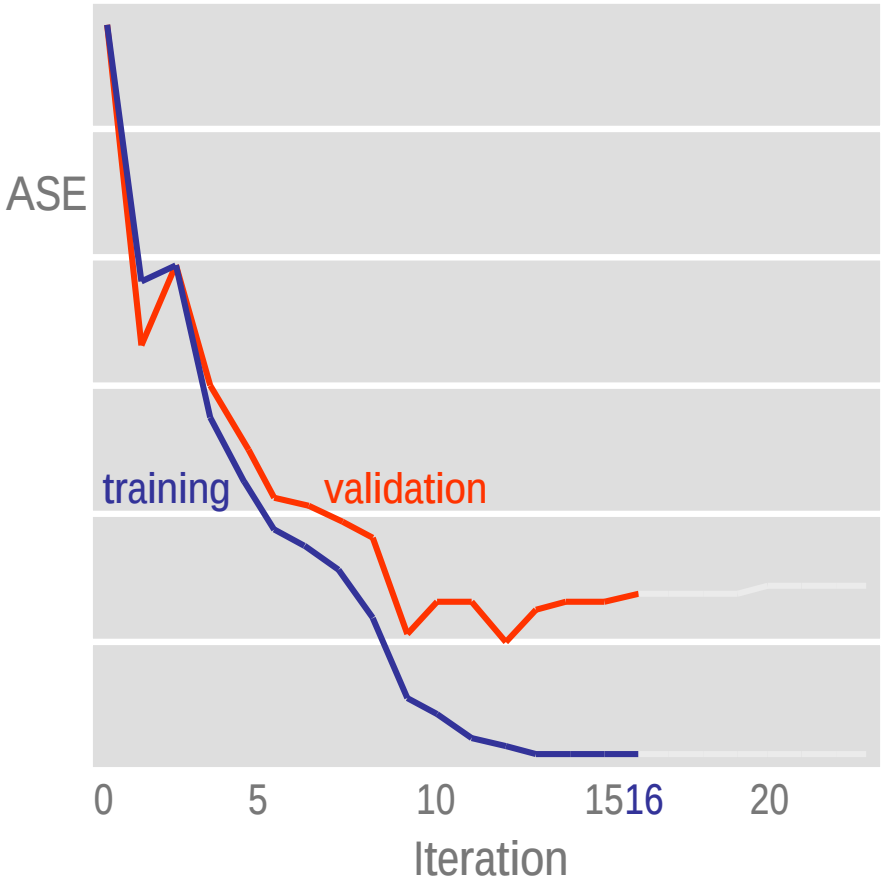
Fit Statistic versus Optimization Iteration

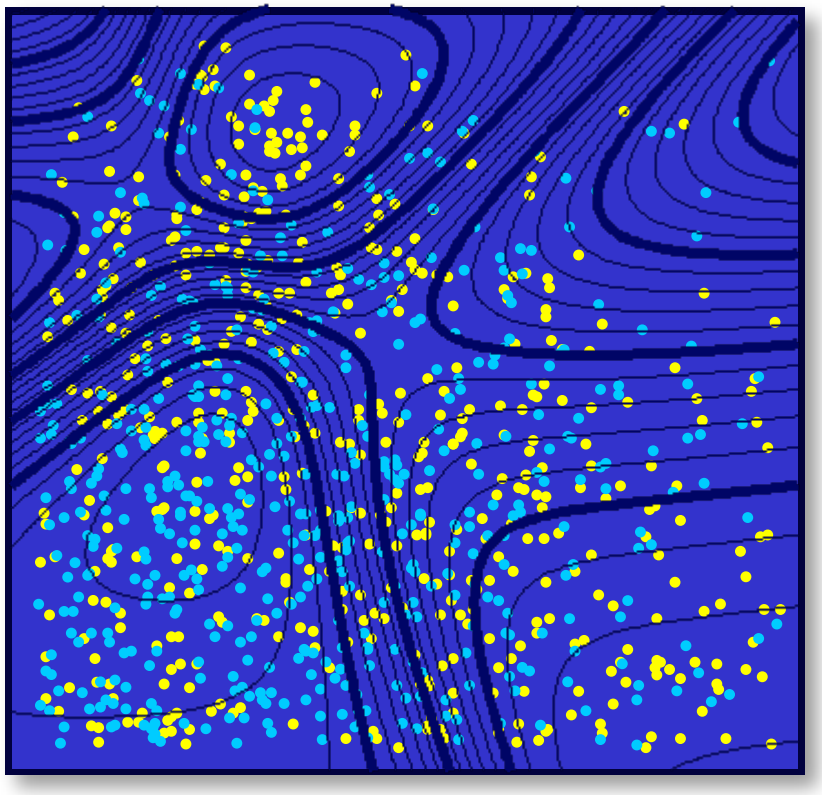




...

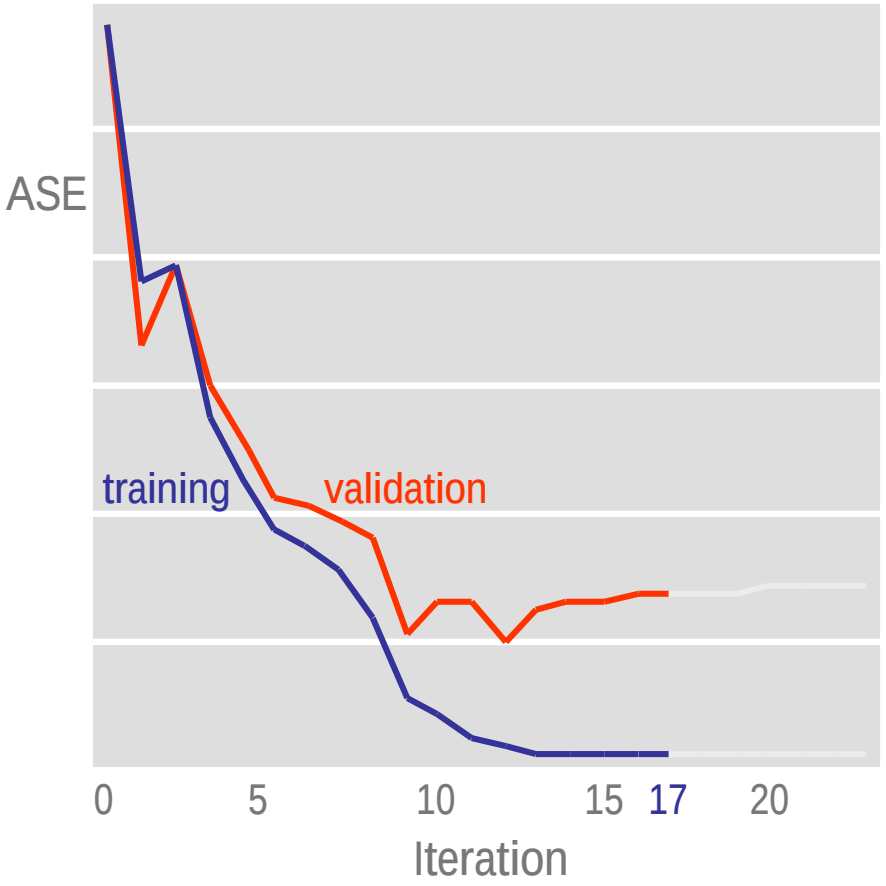
Fit Statistic versus Optimization Iteration

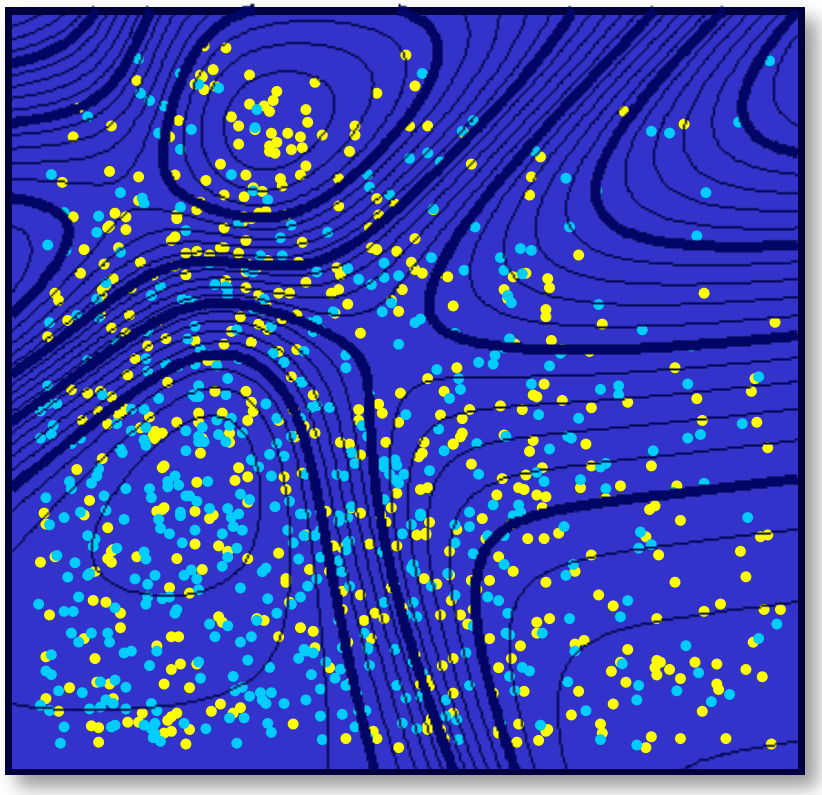




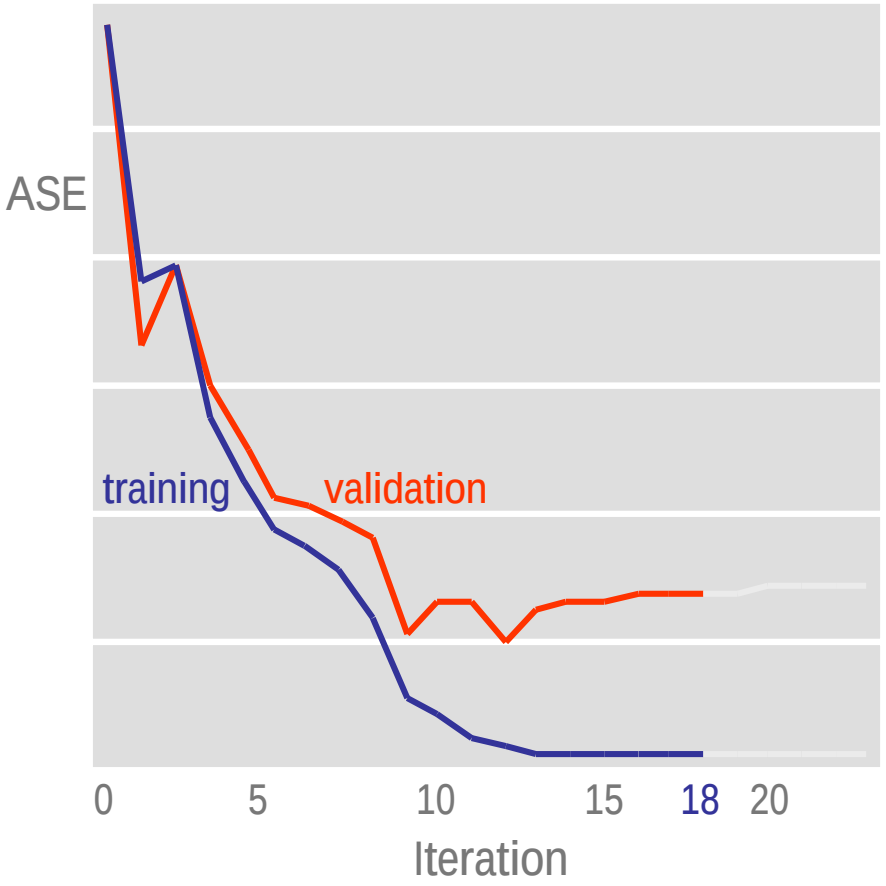
...

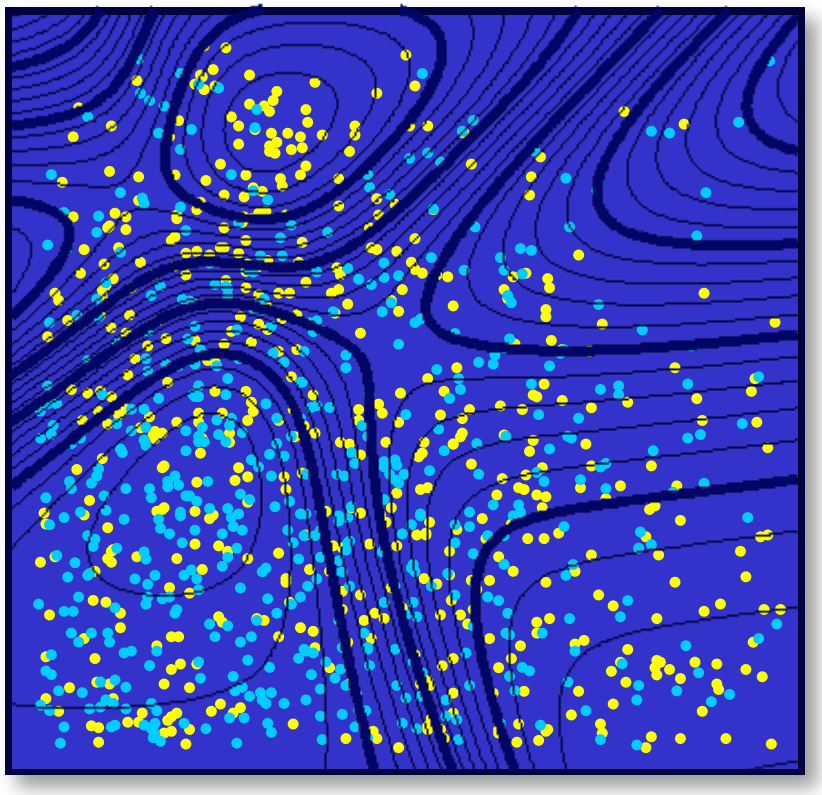
Fit Statistic versus Optimization Iteration



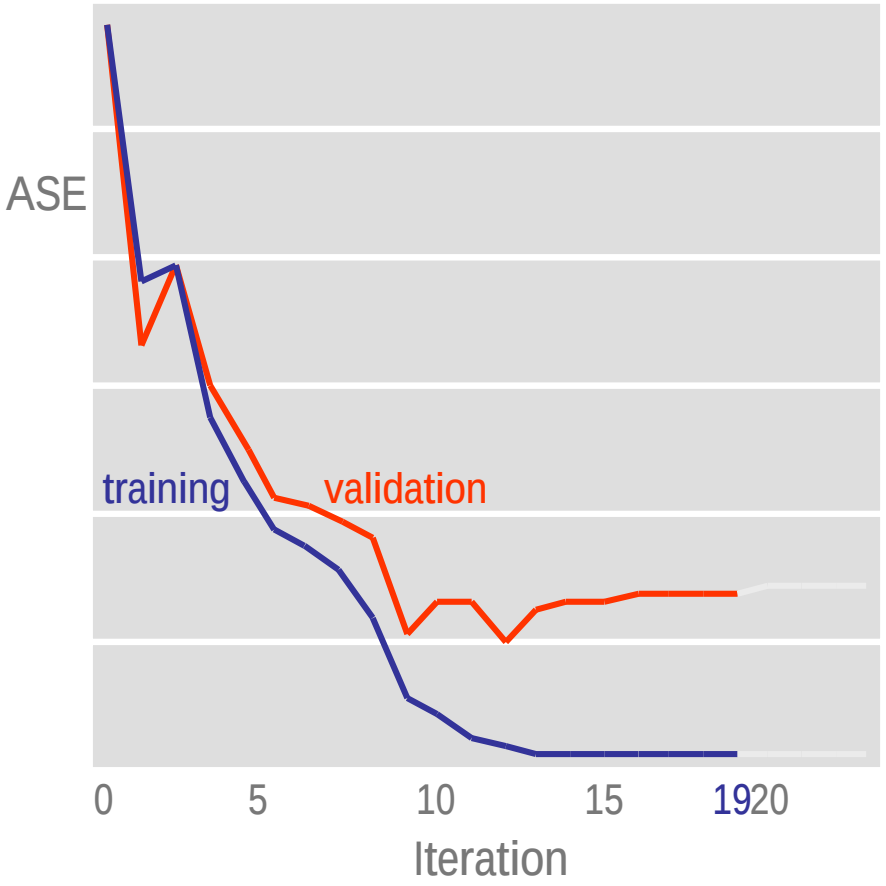


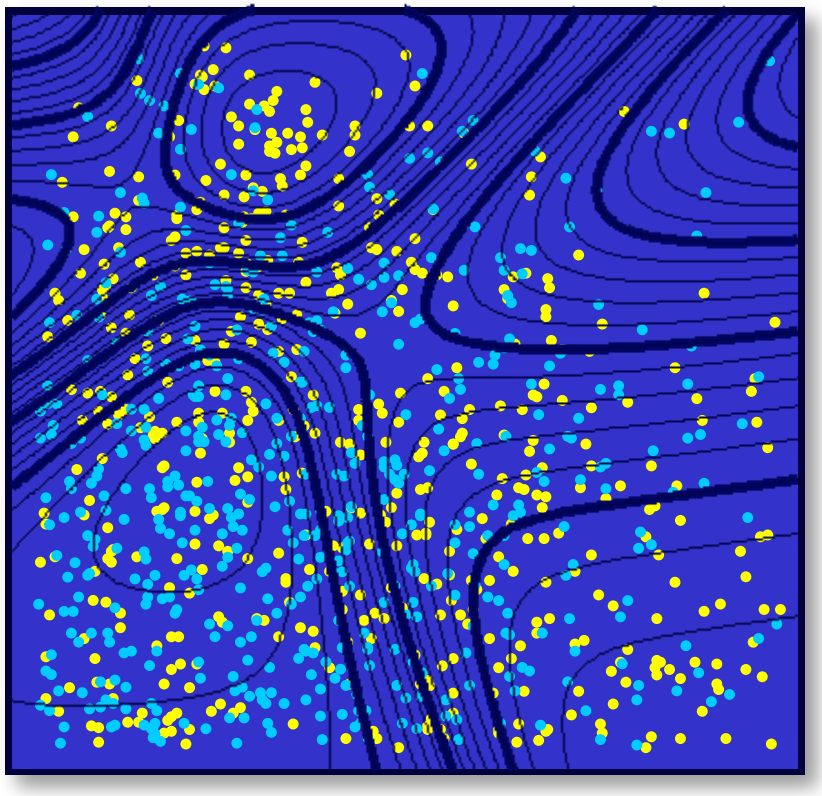
Fit Statistic versus Optimization Iteration





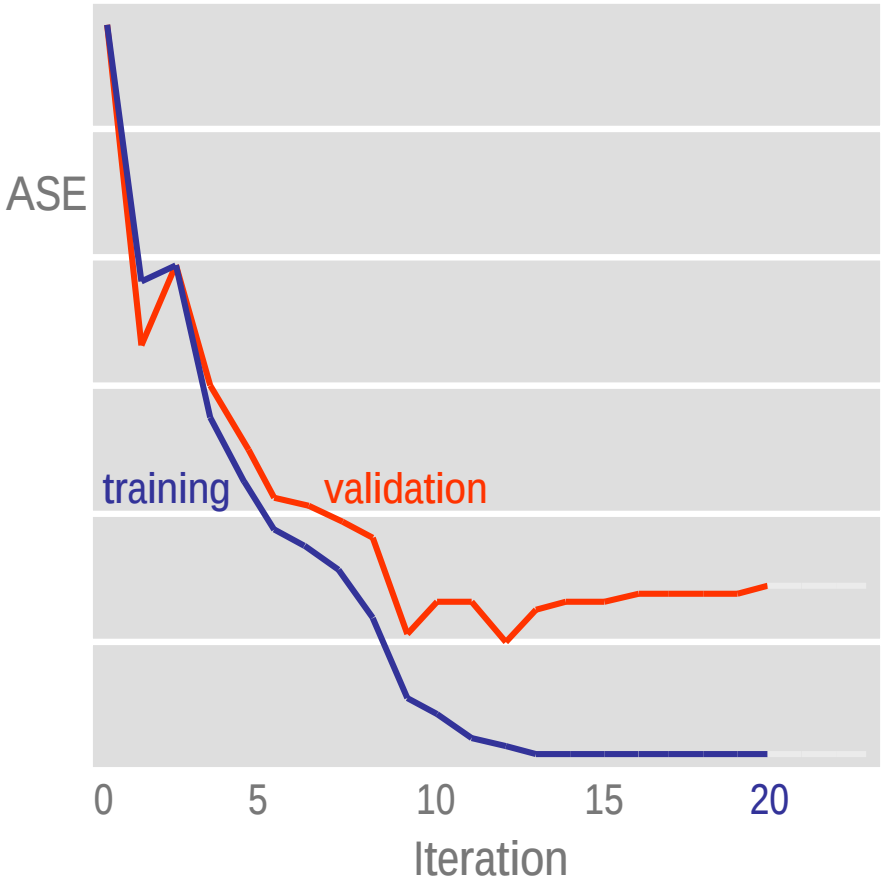
Fit Statistic versus Optimization Iteration

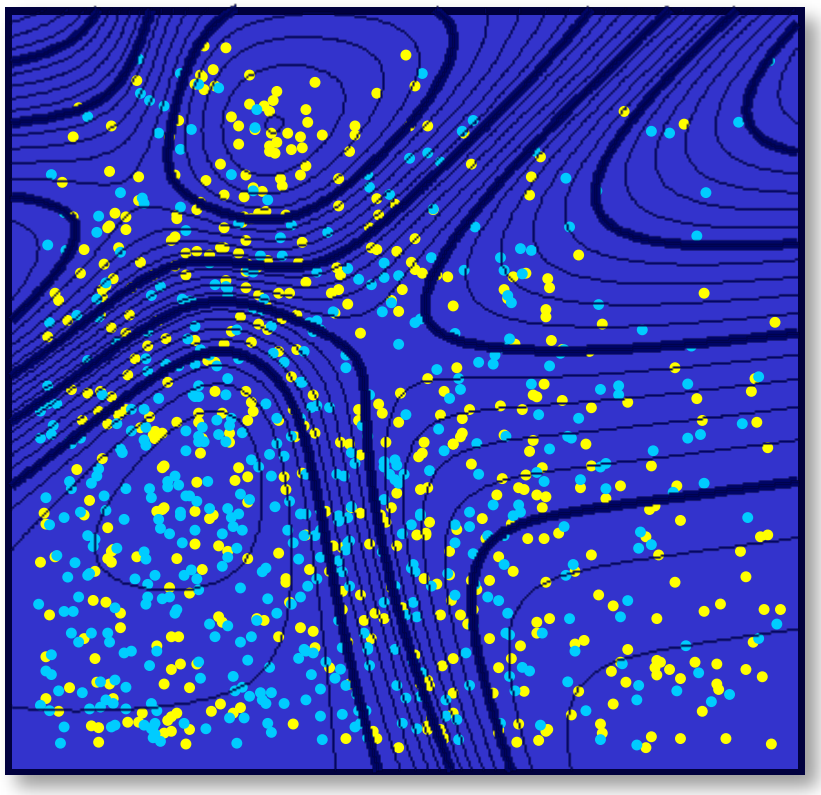




...

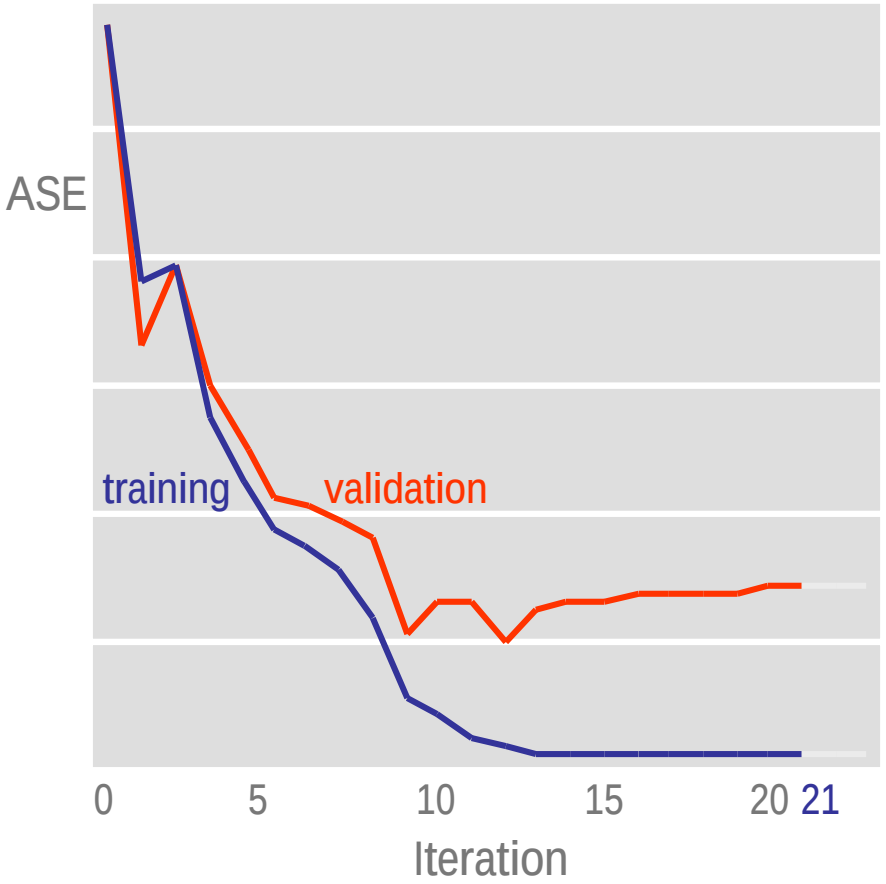
Fit Statistic versus Optimization Iteration

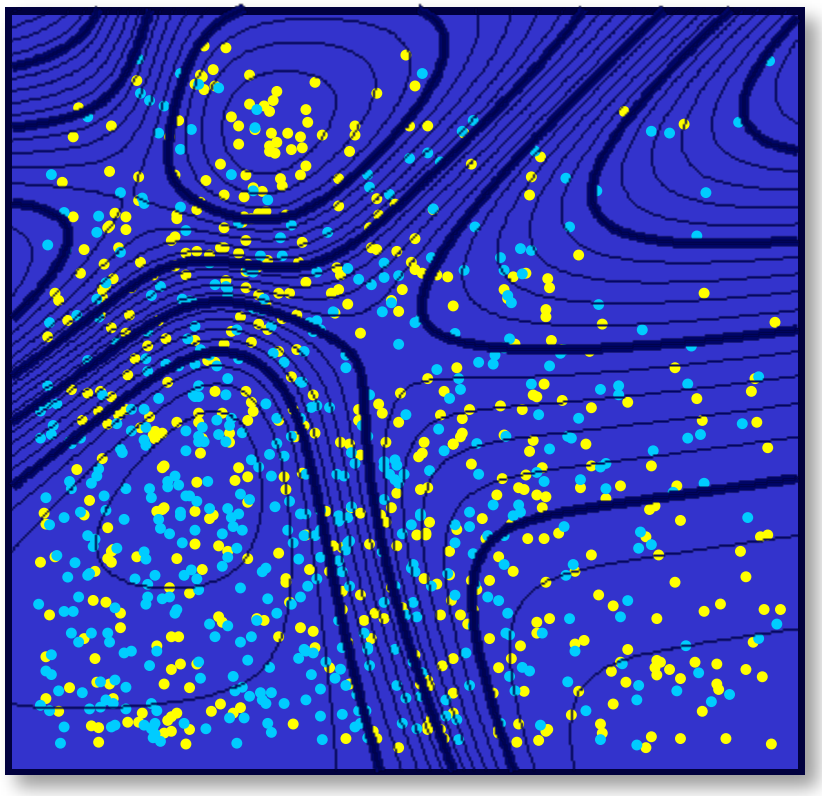




...

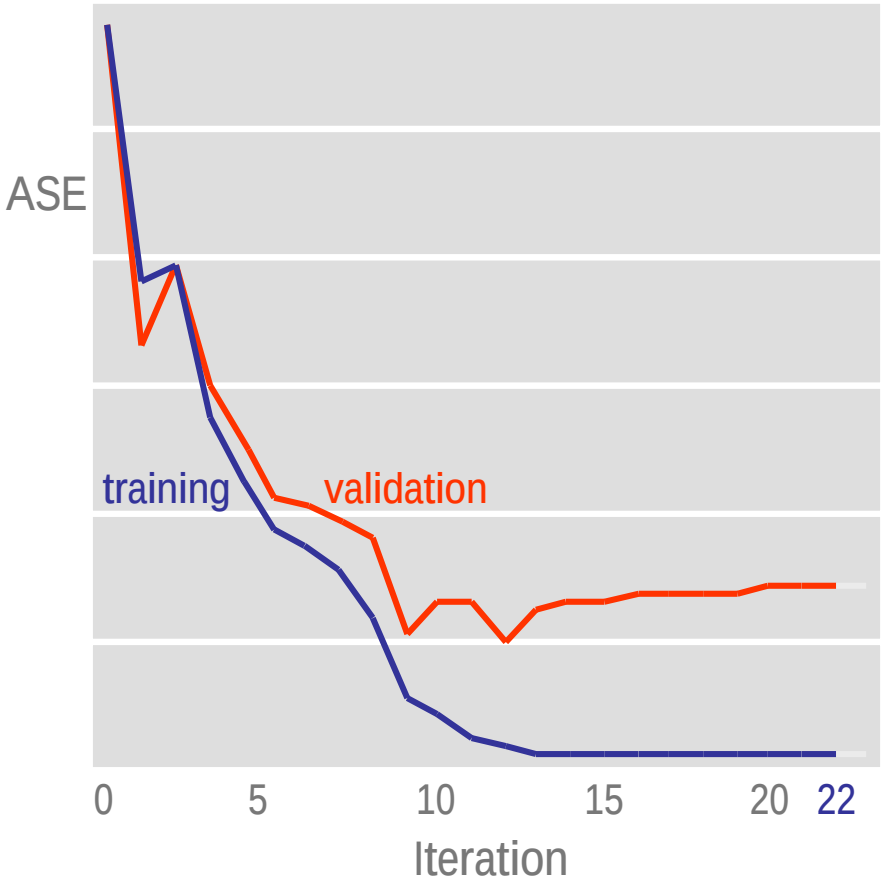
Fit Statistic versus Optimization Iteration

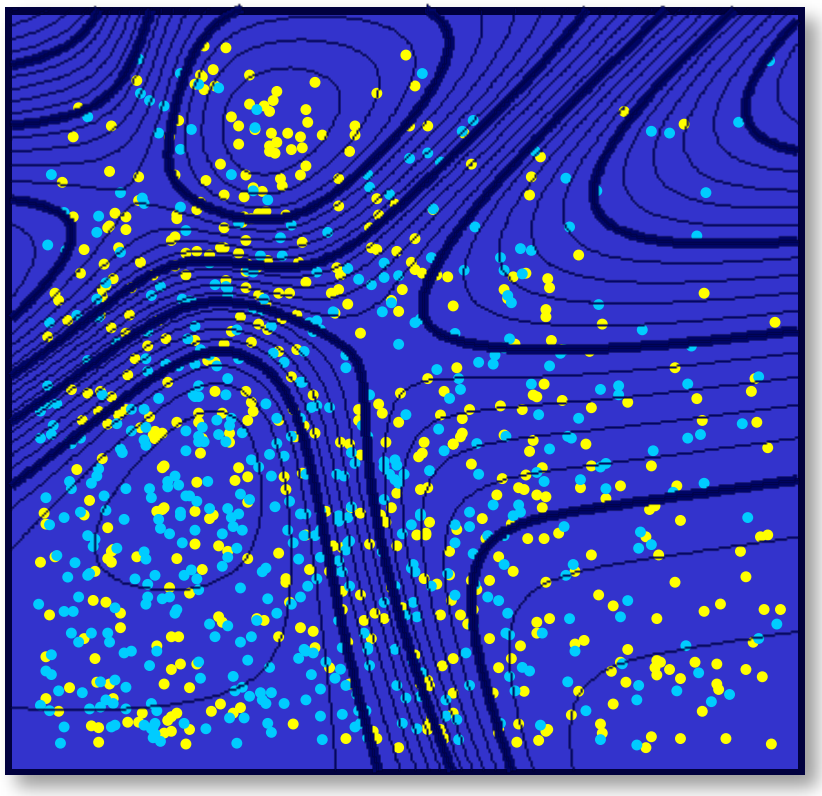




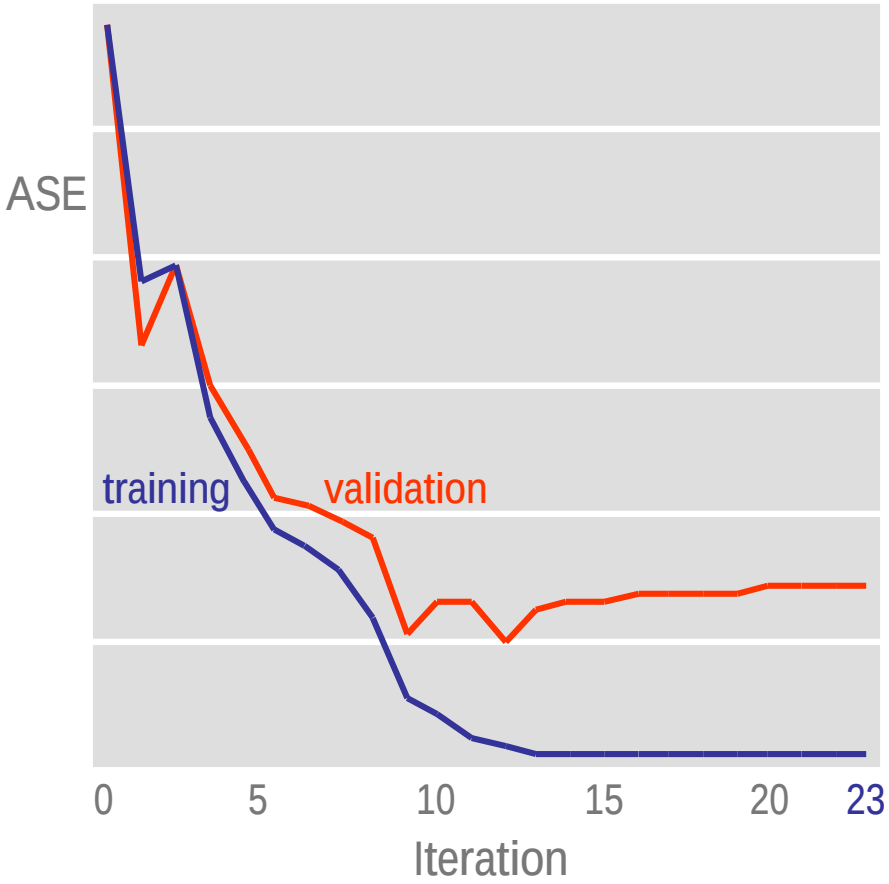
...

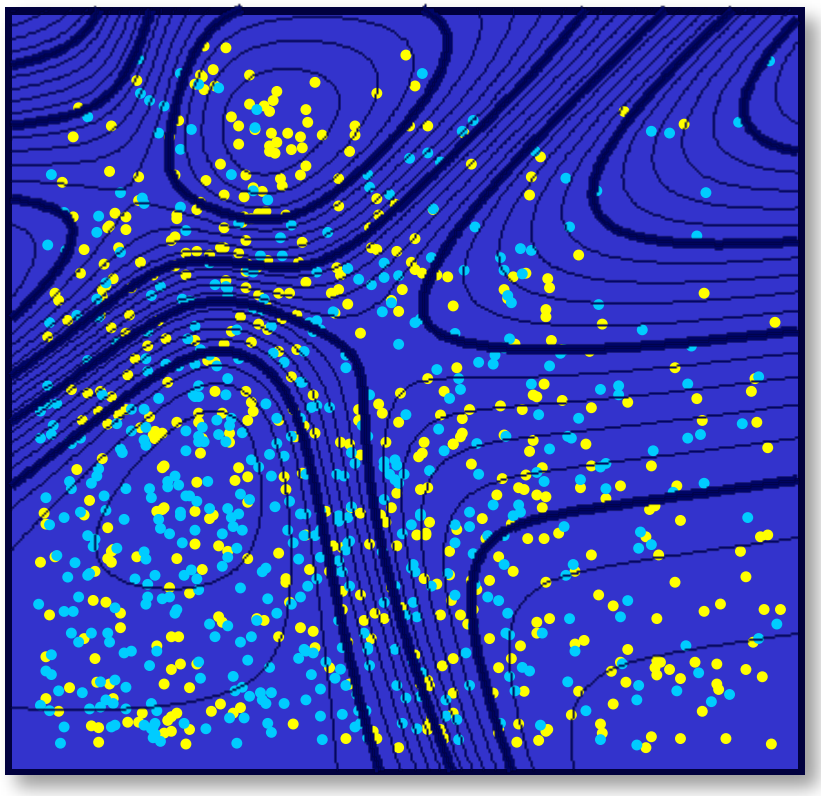
Fit Statistic versus Optimization Iteration





Fit Statistic versus Optimization Iteration





...

Fit Statistic versus Optimization Iteration