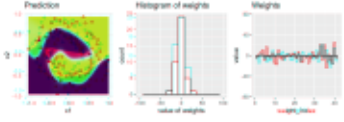
Introduction to Machine Learning

Regularization
Non-Linear Models and Structural Risk Minimization

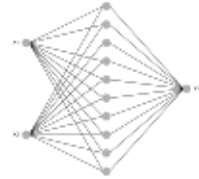


Learning goals

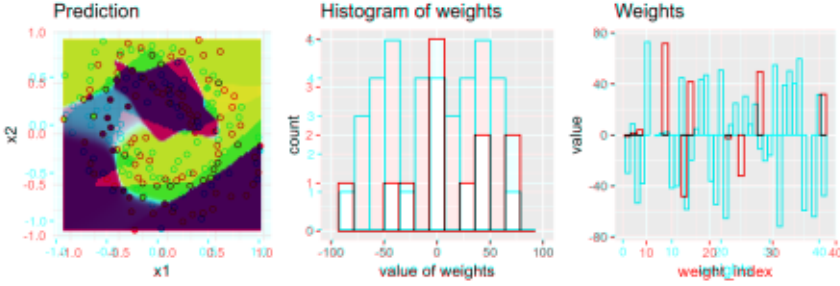
- Regularization even more important in non-linear models
- Norm penalties applied similarly
- Structural risk minimization

REGULARIZATION IN NONLINEAR MODELS

Classification for `spirals` data.
NN with single hidden layer, size 10, $L2$ penalty:



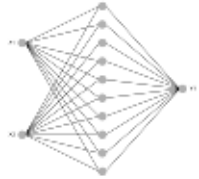
$\lambda = 0$



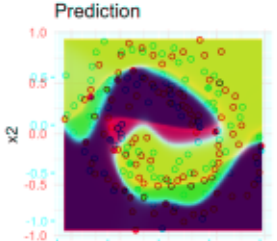
λ affects smoothness of decision boundary and magnitude of weights

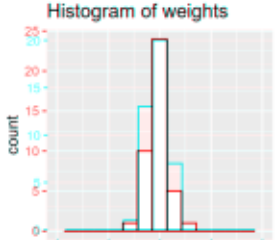
REGULARIZATION IN NONLINEAR MODELS

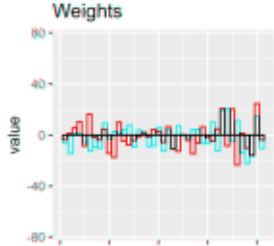
Classification for `spirals` data.
NN with single hidden layer, size 10, $L2$ penalty:



$\lambda = 0.001$



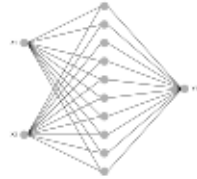




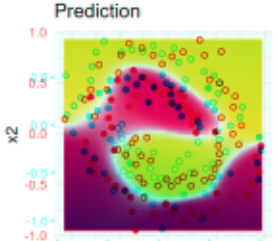
λ affects smoothness of decision boundary and magnitude of weights

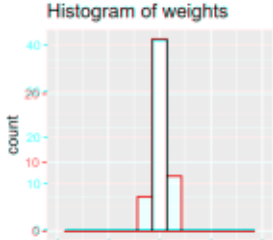
REGULARIZATION IN NONLINEAR MODELS

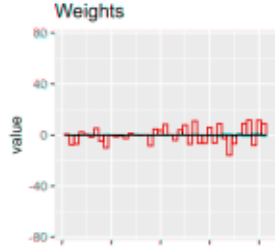
Classification for `spirals` data.
NN with single hidden layer, size 10, $L2$ penalty:



$\lambda = 0.005$



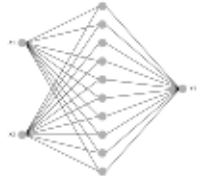




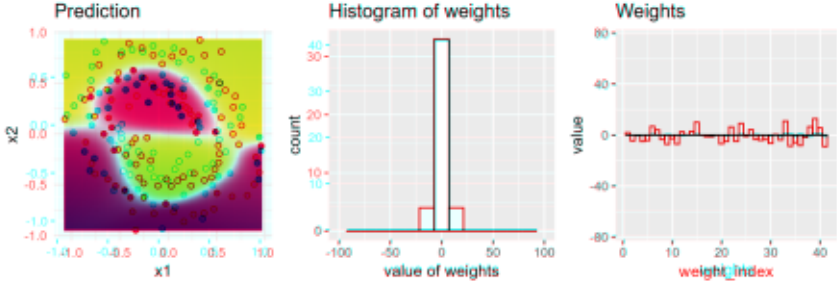
λ affects smoothness of decision boundary and magnitude of weights

REGULARIZATION IN NONLINEAR MODELS

Classification for `spirals` data.
NN with single hidden layer, size 10, $L2$ penalty:



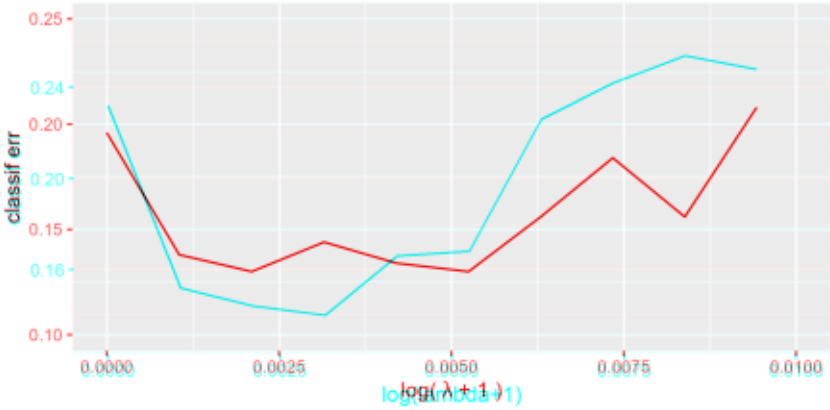
$\lambda = 0.01$



λ affects smoothness of decision boundary and magnitude of weights

REGULARIZATION IN NONLINEAR MODELS

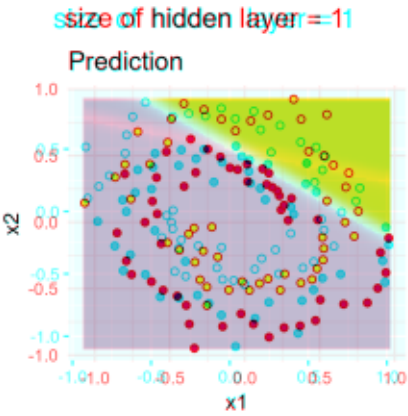
Prevention of overfitting can also be seen in CV.
Same settings as before, but each λ is evaluated with 5x10 REP-CV

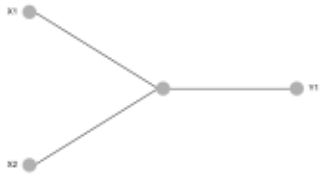


Typical U-shape with sweet spot between overfitting and underfitting

STRUCTURAL RISK MINIMIZATION

Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

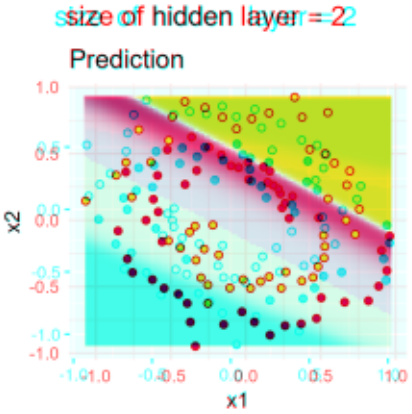


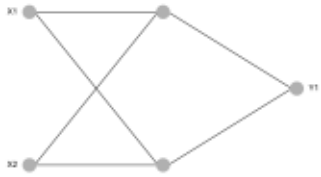


Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

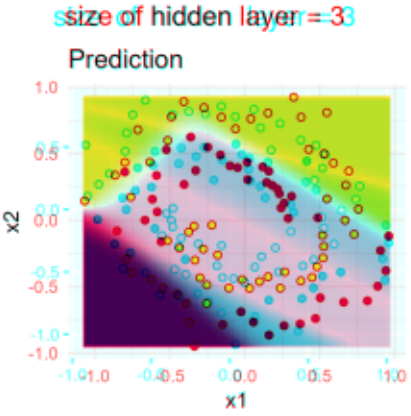


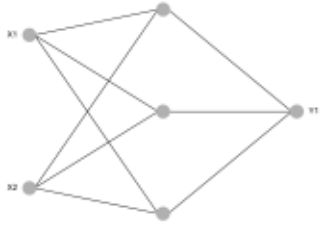


Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

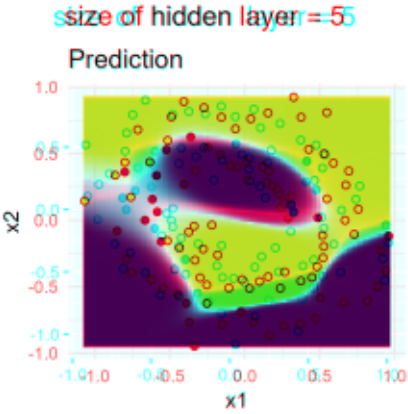




Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

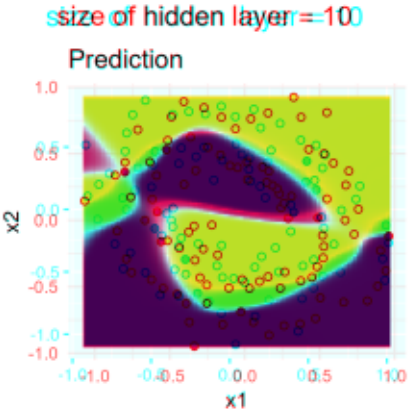
Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

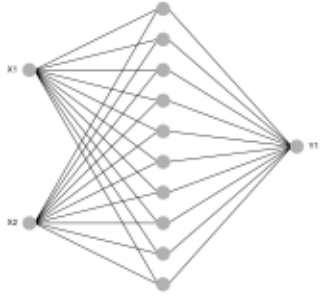


Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

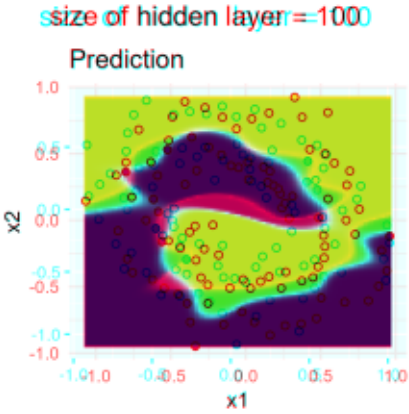


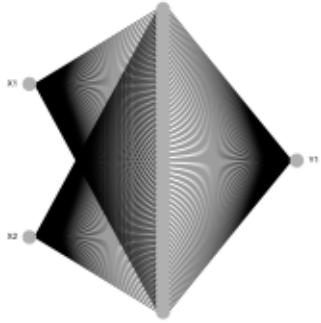


Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

Again `spirals`.
NN with 1 hidden layer, and fixed (small) L2 penalty.

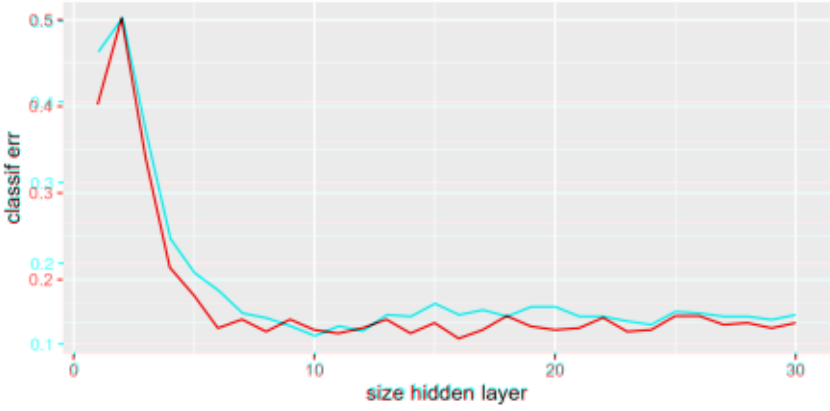




Size affects complexity and smoothness of decision boundary

STRUCTURAL RISK MINIMIZATION

Again, complexity vs CV score.



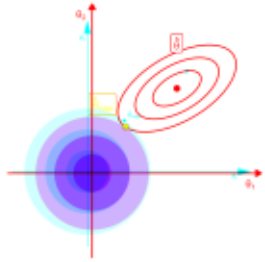
Minimal model with good generalization seems to size=10

STRUCTURAL RISK MINIMIZATION AND RRM

RRM can also be interpreted through SRM,
if we rewrite it in constrained form:

$$\min_{\theta} \quad \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)} \mid \theta\right)\right)$$

$$\text{s.t.} \quad \|\theta\|_2^2 \leq t$$



Can interpret going through λ from large to small as through t from small to large. Constructs series of ERM problems with hypothesis spaces $\mathcal{H}_\lambda$, where we constrain norm of θ to unit balls of growing sizes.