What is the influence of stochastic noise in the dataset on the in- and out-of-sample error of a neural network and how does weight decay regularisation counter this?

The models

Neural network

- Predict $h(x) = P(y = 1|x)$
- Linear transformations with non-linear differentiable ReLU activation functions
- Maximise likelihood by minimising the cross-entropy error
- $(20, 20)$ hidden layers & 2000 epochs

Augmented error regularisation:

$$E_{aug}(h, \lambda, \Omega) = E_{in}(h) + \frac{\lambda}{N}\Omega(h)$$

Weight-decay with L_2 norm:

$$E_{aug}(w) = E_{in}(w) + \frac{\lambda}{N}||w||^2$$

Experimental setup

Repeat for every combination of dataset type, label and data noise:

1. Generate 13 datasets with noise $\in [0, 0.75]$
2. Create 13 models of size 100 with regularisation $\in [0, 1.5]$
3. Train every model on every dataset
4. Compare decision boundaries and training and testing accuracies

Moon dataset with data noise

Conclusions

Support vector machine kernel comparison

How do the linear kernel, polynomial kernel and radial basis function compare to each other, when applied to a synthetic two-dimensional dataset?

Decision tree and k-nearest neighbours regressor forecasting

How does the decision tree regressor model compare to the k-nearest neighbour regressor model in terms of in- and out-of-sample error for time series forecasting?