Machine Learning in Geophysics

Lecture 10 – Dropout, Regularization and other architectures

Overfitting

As with other ML methods we need to be careful to avoid overfitting of training data.

Question

What do we mean by overfitting? How can we idenitify it?

Overfitting

As with other ML methods we need to be careful to avoid overfitting of training data.

Question

What do we mean by overfitting? How can we idenitify it?

Answer

- Training starts to fit noise in the data
- Match between training data and prediction unrealistic
- Significant discrepancy between test and training data performance

Question

How did we avoid overfitting of training data in other ML methods?

Question

How did we avoid overfitting of training data in other ML methods?

Answer

Add Regularization, i.e. limit the variability of the ML parameters. For example, soft-margin SVM

$$\|\mathbf{w}\| + C\sum_{i=1}^{M} \zeta_i \text{ subject to } y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq 1 - \zeta_i.$$

Regularization

We can add regularization to Neural networks

- L_2 regularization, i.e. minimize $\|\mathbf{w}\|_2 = \mathbf{w}^T\mathbf{w}$
- L_1 regularization, minimize $\|\mathbf{w}\|_1 = \sum_i |w_i|$
- Dropout, deactivate random neurons during training

Dropout

- Dropout rates between 0.1 to 0.5
- In each training step different random neurons are not considered
- Neural network is different, some similarity to ensemble methods
- Dropout decreases convergence but can increase robustness
- Validation is performed with full network

Dropout caveats

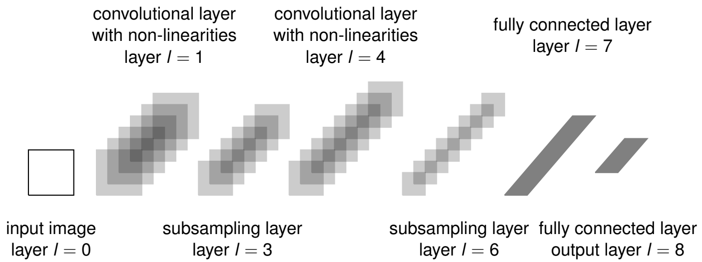
- Average number of connections between training and validation/use differs, needs to be reflected in weights (automatic in Keras)
- SELU activation needs special dropout, Alpha-Dropout, to maintain statistical properties

Monte-Carlo-Dropout

Can further exploit similarity to ensemble methods by averaging predictions of different realizations with dropout.

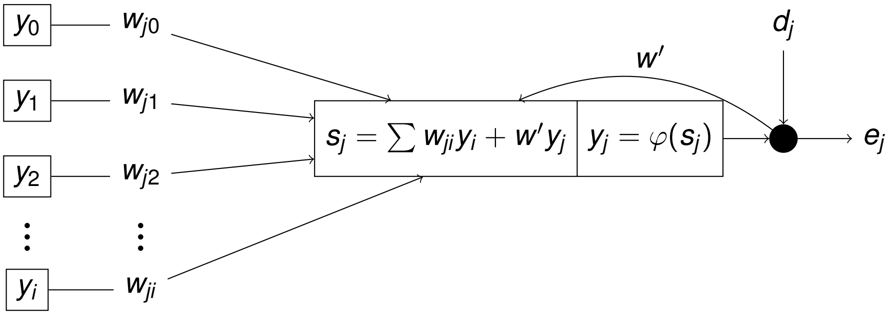
- Generate N realizations of predictions for test data with random dropout
- Average predictions over different realizations
- Can also calculate standard deviations to estimate quality of predictions

Convolutional neural networks



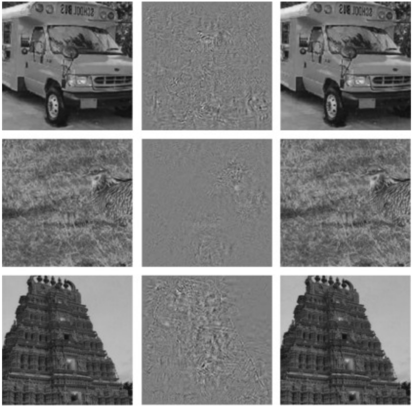
- Each neuron performs convolution with weights instead of simple scalar product
- Convolution closely related to filtering
- Popular in image and audio processing

Recurrent neural networks



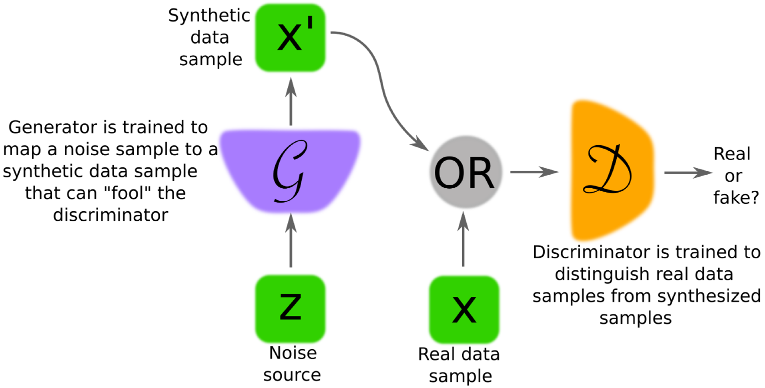
- Output of neurons is fed-back to its own input
- Provides network with memory
- Used for time-series analysis and prediction

Generative adversarial networks - GAN



- Small changes can confuse NN discriminators
- Try to generate such changes to improve training

Generative adversarial networks - GAN



- Generator will become better at creating realistic representations

Summary

- ML can be used for a wide range of classification and regression tasks
- Data driven, need good training data (for supervised learning)
- Large number of architectures, parameters, approaches, often trial-and-error
- Need clear questions and suitable data
- Overfitting can suggest great results but means poor general performance
- Need to critically evaluate our results