Neural Networks

1 Questions

1. What importance does the database have for the success of the training process? What requirements should a training database fulfill?
2. What is meant by the term "overfitting"? What countermeasure is usually taken to prevent this?
3. Give possible reasons why neural networks are often able to yield a better performance than other statistical methods.
4. Why are extended vectors introduced (slide 30)? What advantage does the introduction of the "helping" variable (slide 40) have? Why do network activation functions have to be differentiable for this approach?
5. What is the role of the "random generation" block (slide 47) in the GAN training process? Why is the squared classification error of the discriminator network at one point minimized and at another point maximized?

2 Answers

1. The database is the only source of knowledge and information used in the training process. Therefore, the contents of the database need to be carefully chosen. Since a network cannot learn to deal with cases it was never presented with during training, sophisticated topologies (e.g. c onvolutional) will significantly underperform without the proper database. A "proper" database should include all possible data variations (e.g. scaling, translation, rotation) if no separate preprocessing is applied. It also needs to include a "large enough" number of case realizations to be able to adapt all free parameters of the network.

2. An overfitted network lacks the ability of generalization. While, based on a cost function, the network performs well for training data, performance for any unseen testing data will be poor. The most rudiment method to counteract overfitting is to split the database in separate training / evaluation / testing sets. During training runtime, the evaluation data is used to check for convergence and for overfitting. The testing data is used after training to determine the actual network performance.

3. Due to the different topologies and the "deep" character, complex multivariate relations can be established. The activation functions also allow for nonlinearities.

4. Vectors are extended to yield the activation function input value via a single vector multiplication . By doing so, all adaptable parameters connecting into one activation function are included in the same vector and the update functions can be derived more easily. The parameter update becomes easier to read and may be written and implemented in an iterative way (slide 45). The update of the helping variable requires the derivative of the activation function . This is due to the concept of gradient descent which applies partial derivatives.

5. The random generator "decides" whether the discriminator network "sees" the desired data or its estimate. The minimization is required for the training of the discriminator network, since the network's task is to perform the classification true data vs. estimate. For the training of the generator network, however, the classification error should be maximized to make the distinction between true data and its estimate "impossible".