WOLFRAM   PRODUCTS   SOLUTIONS   PURCHASE   SUPPORT   COMMUNITY   COMPANY   OUR SITES   SEARCH

Wolfram Mathematica
DOCUMENTATION CENTER

SEARCH MATHEMATICA 8 DOCUMENTATION

›

NEURAL NETWORKS DOCUMENTATION   ◀ Previous | Next ▶

2.8 Unsupervised and Vector Quantization (VQ) Networks

Unsupervised ?algorithms are used to find structures in the data. They can, for instance, be used to find clusters of data points, or to find a one-dimensional relation in the data. If such a structure exists, it can be used to describe the data in a more compact way.

Most network models in the package are trained with *supervised* training algorithms. This means that the desired output must be available for each input vector used in the training. *Unsupervised* networks, or *self-organizing networks*, rely only on input data and try to find structures in the input data space. The training algorithms are therefore called *unsupervised*.

Since there is no "correct" output, there will also not be any "incorrect" outputs. This fact leaves a lot of responsibility to the user. After an unsupervised network has been trained, it must be tested to show that it makes sense, that is, if the obtained structure is really representing the data. This validation can be very tricky, especially if you work in a high- dimensional space. In two- or three-dimensional problems, you can always plot the data and the obtained structure and simply examine them. Another test that can be applied in any number of dimensions is to check for the mean distance between the data points and the obtained cluster centers. A small mean distance means that the data is well represented by the clusters.

An unsupervised network consists of a number of *codebook vectors,* which constitute cluster centers. The codebook vectors are of the same dimension as the input space, and their components are the parameters of the unsupervised network. The codebook vectors are called the *neurons* of the unsupervised network.

When an unsupervised network is trained, the locations of the codebook vectors are adapted so that the mean Euclidian distance between each data point and its closest codebook vector is minimized. The algorithm, called competitive learning, is described in Section 10.1.2, UnsupervisedNetFit.

An unsupervised network can employ a *neighbor feature*. This gives rise to a s*elf-organizing map* (SOM). For SOM networks not only the mean distance between the data and nearest codebook vector is minimized, but also the distance between the codebook vectors. In this way it is possible to define one- or two-dimensional relations among the codebook vectors, and the obtained SOM unsupervised network becomes a nonlinear mapping from the original data space to the one- or two-dimensional feature space defined by the codebook vectors. Self-organizing maps are often called *self-organizing feature maps*, or *Kohonen networks*.

When the data set has been mapped by a SOM to a one- or two-dimensional space, it can be plotted and investigated visually.

The training algorithm using the neighbor feature is described in Section 10.1.2, UnsupervisedNetFit.

Another neural network type that has some similarities to the unsupervised one is the *Vector Quantization* (VQ) network, whose intended use is classification. Like unsupervised networks, the VQ network is based on a set of *codebook vectors*. Each class has a subset of the codebook vectors associated to it, and a *data vector* is classified to be in the class to which the closest codebook vector belongs. In the neural network literature, the codebook vectors are often called the *neurons* of the VQ network.

Each of the codebook vectors has a part of the space "belonging" to it. These subsets form polygons and are called Voronoi cells. In two-dimensional problems you can plot these Voronoi cells.

The positions of the codebook vectors are obtained with a supervised training algorithm, and you have two different ones to choose from. The default one is called *Learning Vector Quantization* (LVQ) and it adjusts the positions of the codebook vectors using both the correct and incorrect classified data. The second training algorithm is the competitive training algorithm, which is also used for unsupervised networks. For VQ networks this training algorithm can be used by considering the data and the codebook vectors of a specific class independently of the rest of the data and the rest of the codebook vectors. In contrast to the unsupervised

networks, the output data indicating the correct class is also necessary. They are used to divide the input data among the different classes.

Any questions about topics on this page?
Click here to get an individual response.

BUY NOW >> MORE INFO >>

About Wolfram | Wolfram Blog | Wolfram|Alpha | Terms | Privacy | Site Map | Contact

Newsletter »