1 Hopfield-like network for image classification

Inspired by the simplicity of Hopfield networks, we consider the idea of using the same kind of Hebbian learning rule in a feed-forward architecture.

Consider a 1-layer network with N input neurons σ_i, P output neurons S_j and synapses ω_{ji}

$$S_j = \sum_i^N \omega_{ji}\sigma_i \,. \tag{1}$$

In the training phase, when a pattern σ_i^1 is presented to the input neurons, it is stored in the synapses connected to the first output neuron. A second pattern σ_i^2 is stored in the synapses connected to the second output neuron and so on:

$$\omega_{i1} = \sigma_i^1 \,, \qquad \omega_{i2} = \sigma_i^2 \,, \qquad \omega_{ij} = \sigma_i^j \,. \tag{2}$$

The network will then be able to store P patterns.

When a new pattern σ_i^k is presented to the input layer, the network response will be stronger at the output neuron S_j corresponding to the stored pattern σ_i^j more similar to σ_i^k. Note that this is true only for binary patterns ($\sigma_i = \pm 1$). To prove this statement, simply consider that the product of 2 neurons σ_i, σ_i' is maximum when they are equal. So the less neurons differ between the input and the stored pattern, the higher the network response will be.

When using continuous input neurons, the network response tends to fall into the "brightest" stored patterns. This is not a limitation however, because it is always possible to replicate the network n times to be able to store n-bit representations of continuous inputs.

The network can be also propagate signals backwards

$$\sigma_i = \sum_j^P \omega_{ij} S_j \,. \tag{3}$$

By firing a single neuron k in the output layer, the response of the network in the input layer will be the corresponding stored pattern σ_i^k

$$S_j = \delta_j^k \longrightarrow \sum_j^P \omega_{ij}\delta_j^k = \omega_{ik} = \sigma_i^k \,. \tag{4}$$

So, as in the original Hopfield model, the network is able to remember the stored patterns.

An advantage of this approach over the standard Hopfield implementation is that the capacity of the network is independent of the input layer dimension. If necessary the output size can be increased to accomodate more patterns without affecting the already stored ones. Also, the interference between stored patterns that is observed in the Hopfield model due to the overlaps in the synaptic weights is avoided in this architecture since each synapses contains information regarding only a single pattern.

It could be interesting to investigate if by combining this approach with other mechanisms it would be possible to obtain an network able to learn unsupervised to segment and differentiate objects.