AG Neuroinformatik
Prof. Dr. Helge Ritter

Universität Bielefeld
Technische Fakultät

Advanced Machine Learning

Problem Sheet 2
due Friday, Nov-20 2015

Problem1: (echo state network) Implement an echo state network described by the following dynamics:

$$s_i(t) = (1-\lambda)s_i(t-1) + \lambda tanh\left(\sum_{j=1}^{N} a_{ij}s_j(t-1) + b_i x(t)\right)$$

$$y(t) = \sum_{i=1}^{N} w_i s_i(t)$$

where s_i are the neuron activities, and x/y are scalar inputs/outputs.

Create a network of 100 neurons with random connections a_{ij} and b_j drawn from a normal distribution $N(0,\sigma_a)$, $N(0,\sigma_b)$ where σ_a, σ_b parametrize the spread of the input and the inner weights (for the following, it is advantageous to visualize the network outputs as time series plots, ideally with an interactive possibility to explore parameter changes)

a) consider the dynamics of the network in the absence of any inputs for $\lambda = 0.2$. Explain why in this case σ_b is irrelevant, and explore the behavior of the network for different values of σ_a. How does the network activity (plot the neuron activities as gray levels of a 10x10 matrix; also plot the activities of two arbitrary neurons in a 2d scatter diagram) differ for small vs. large values of σ_a, and approximately where does the transition between both regimes occur (Hint: since each new random "ruins" the current state, the transition behavior is more clearly seen by scaling a fixed random weight matrix without destroying the state information). Explain the consequences w.r.t. using the network as a dynamic reservoir network!

b) Given a time sequence of desired target outputs y^{target}, consider the following on-line adaptation rule for the output weights w_i of the network:

$$\Delta w_i = \epsilon \left(y_t^{\text{target}} - y_t\right) s_i(t)/||s(t)||^2$$

Explain the rationale behind this rule!

c) Explore whether the above learning rule can be used to train the reservoir (without any input) to generate an approximately sinusoidal wave pattern. Which values of σ_a will you need, and which frequencies can be realized best?

d) Finally study training of the network to predict the next time step of a triangle wave with period of 10 time steps. Which σ_a works best here?

Problem 2 (Gaussian process interpolation): consider the representation of input-output data $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{2N}$ with a Gaussian process:

$$f(x) = \boldsymbol{k}(x)^T (\boldsymbol{K} + \sigma^2 \mathbf{1})^{-1} \boldsymbol{y}, \qquad \text{where} \tag{1a}$$

$$K_{ij} = \exp(-|x_i - x_j|^2 / L) \qquad \text{and} \tag{1b}$$

$$k_i(x) = \exp(-|x - x_i|^2 / L). \tag{1c}$$

Parameter L specifies the correlation length of the Gaussian process, σ is the standard deviation per element for the data $\boldsymbol{y}$.

a) show that Eq. (1) for $\sigma = 0$ and arbitrary data tuples $(\boldsymbol{x}, \boldsymbol{y})$ leads to $f(x_i) = y_i$, i.e. the graph defined by [Gauss:1] passes through all specified data points!

b) implement eq. (1) as algorithm in a programming language of your choice (e.g. Octave/Matlab/Python) and plot the resulting graphs for the data $\boldsymbol{x} = (0, 1, 3, 4)$ and $\boldsymbol{y} = (0, 3, 3, 0)$ and parameter $L = 2, \sigma = 0.1$. Which function value $f(x)$ would you expect at $x = 2$ and what is delivered by the Gaussian process?

c) Compute values $a = f(x = 1.1)$ and $b = f(x = 1.2)$ and add now as further data points $(1.1, a + \epsilon)$ and $(1.2, b)$ and plot the function graph for values $\epsilon = 0$, $\epsilon = -0.1$ and $\epsilon = 0.1$. Discuss the observed behavior! What happens for larger σ ?