equations_doc

May 28, 2020

1 Documenting the Echo State Network equations

1.1 Introduction

An echo state network (ESN) is a special kind of neural network made of a large reservoir of N neurons. That's why some people refer to echo state networks as reservoir computing methods. Among the different echo state networks, the one we will explain in this doc is a speciffic one whose main goal is to predict the continuation of a signal.

1.2 The equations

Our reservoir of neurons is described by a state $\mathbf{x} \in \mathcal{R}^N$ that evolves when receiving a feedback value $y_{fb} \in \mathcal{R}$. The value y_{fb} can either come from the time series of a signal we want to learn or from a value that the ESN itself has generated. The equation that updates a reservoir state $\mathbf{x}(n)$ to a new state $\mathbf{x}(n+1)$ is the following:

$$\widehat{\mathbf{x}}(n+1) = \tanh(u_{in}\mathbf{w_{in}} + W_{res}\mathbf{x}(n) + \mathbf{w_{fb}}y_{fb}(n))$$

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \alpha(\widehat{\mathbf{x}}(n+1) - \mathbf{x}(n))$$

The vector $\mathbf{w_{in}} \in \mathcal{R}^N$, the matrix $W_{res} \in \mathcal{R}^{N\times N}$ and the feedback vector $\mathbf{w_{fb}} \in \mathcal{R}^N$ come from a uniform random distribution between $[-1, 1)$. All these random parameters will not be changed, once initialized they will remain fixed. The constant value $u_{in} \in \mathcal{R}$ helps to provide numerical stability and can take a small value like 0.1. It also appears in the ridge regression equation that we will explain later. The matrix W_{res} must be very sparse and it must have a spectral radius ρ_{spec} lower than 1. The program chooses $\rho_{spec} = 0.8$ and a sparsity of 10 connections per neuron (only 10 non zero values per row). Finally, the value α is called the leaking rate and it can be close to 1 (the program selects 0.9). The main goal of alpha is to provide stability to the update state equation.

1.3 The teacher forcing process

The teacher forcing process has the goal to learn a signal in order to be able to predict its continuation later. In this process we feed the ESN with the time series y_{signal} we want to learn. The teacher forcing process starts by running the update state equation as many times as points we have in our time series. If our time series has m points $\{y_{signal}(1), y_{signal}(2)...y_{signal}(m)\}$, we can initialize $\mathbf{x}(1) = 0$ and generate a collection of m+1 reservoir states $\{\mathbf{x}(1), \mathbf{x}(2), ...\mathbf{x}(m+1)\}$. Note that the final state $\mathbf{x}(m+1)$ is generated with the last pair $\{\mathbf{x}(m),\ y_{signal}(m)\}$. Once this data has been generated, a ridge regression problem must be solved:

$$\mathbf{w_{out}}, b = argmin\left(\sum_{i=s}^{m}\left(y_{signal}(i) - y_{pred}(i)\right)\right)^2 + \beta(||\mathbf{w}_{out}||_2^2 + b^2)\Bigg)$$

Where $y_{pred} =< \mathbf{w_{out}}, \mathbf{x}(i) > +u_{in}b$

We have used <.> to denote the scalar product and $||.||_2$ to denote the second norm. The parameter β is a regularization parameter that controls how much we penalize the second norm of the solution $\mathbf{w_{out}}$, b. The integer s lies in the range $[1, m)$ and excludes from the equation the initial states $\mathbf{x}(i < s)$ that suffer a transcient period due to $\mathbf{x}(1) = 0$. In the program, the variable s takes the name num_skip.

1.3.1 About regularization

The regularization parameter β makes the weights $\mathbf{w_{out}}$ and b take small values, which helps the ESN to generalize and make a prediction that is more stable. It also exists a simpler regularization technique, which consists in adding small random values (noise) to the training signal while keeping $\beta = 0$. In some cases, for example when learning a signal called the Mackey glass, this technique leads to much better predictions that the ones you can obtain with different values of β.

1.4 The Prediction

The ESN runs the update state equation receiving feedback from the values $\{y_{pred}(m+1), y_{pred}(m+2), ...)\}$ that are being predicted with $y_{pred}(i) =< \mathbf{w_{out}}, \mathbf{x}(i) > +u_{in}b$.