VC Dimension Bounds for Networks of Spiking Neurons

Michael Schmitt

Lehrstuhl Mathematik und Informatik, Fakultät für Mathematik
Ruhr-Universität Bochum, D-44780 Bochum, Germany
`mschmitt@lmi.ruhr-uni-bochum.de`

Abstract. We calculate bounds on the VC dimension and pseudo dimension for networks of spiking neurons. The connections between network nodes are parameterized by transmission delays and synaptic weights. We provide bounds in terms of network depth and number of connections that are almost linear. For networks with few layers this yields better bounds than previously established results for networks of unrestricted depth.

1. Introduction and Definitions

Due to recent neurobiological findings it has become increasingly clear that computation in natural neural systems is not based solely on average firing rates but also on the timing of single spikes. In such networks parameters are important that are not captured by conventional neuron models such as threshold or sigmoidal gates. In this paper we study a model for networks of spiking neurons where the connections between nodes are parameterized by synaptic transmission delays. Recent theoretical results have shown that the computational power and learning capabilities for a spiking neuron with adjustable delays are significantly higher compared to neuron models that only have weights as programmable parameters [4, 7]. Moreover, by exhibiting networks with quadratic VC dimension the number of training examples that an algorithm needs for adapting the delays of a network of spiking neurons has been shown to grow at least quadratically in the number of adjustable delays [5]. (For details concerning the relationship between VC dimension and learnability we refer the reader to Vidyasagar [8].) This result, however, uses networks of unrestricted depth. Here, we analyze the VC dimension and pseudo dimension for networks of spiking neurons in terms of the depth (or number of layers) and the number of connections of the network. We provide upper bounds that are almost linear in these parameters. Further, by establishing linear lower bounds we show that these results are almost optimal.

In a network of *spiking neurons* each node v receives inputs in the form of short pulses, or *spikes*, through its incoming connections from other nodes. Each connection is parameterized by two numbers: its weight $w_i \in \mathbb{R}$ and its delay $d_i \in \mathbb{R}^+$ (where $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$). We assume that if v receives a spike through its i-th connection which has been emitted by some node at time t_i, this generates a rectangular pulse in v described by $h_{d_i,w_i}(t - t_i)$ with

$$h_{d_i,w_i}(t) = \begin{cases} 0 & \text{for } t < d_i \text{ or } t \geq d_i + 1 \,, \\ w_i & \text{for } d_i \leq t < d_i + 1 \end{cases}$$

where we treat time as a continuous variable denoted by t. The *membrane potential* of v is a superposition of these pulses defined by $P_v(t) = \sum_{i \in F} h_{d_i,w_i}(t - t_i)$ where F is the set of those connections through which v receives a spike. Node v emits a spike as soon as P_v reaches a certain threshold θ_v. More precisely, the *firing time* t_v of v satisfies $t_v = \min\{t : P_v(t) \geq \theta_v\}$.

The neuron model that we consider here is a simple version of a *leaky integrate-and-fire neuron* and has been introduced by Maass [2]. The functions h_{d_i,w_i} approximate the so-called postsynaptic potentials of biological neurons and are used for implementations of pulsed neural networks in analog VLSI such as described by [6]. Further discussions of this type of neuron model can be found in the surveys by Gerstner [1] and Maass [3].

We consider computations of spiking neurons based on two types of coding. *Binary coding* is used to represent Boolean values assuming that a neuron fires at some fixed time if it encodes a 1, and that it does not fire at all if it encodes a 0. For *analog coding* the firing of a neuron at time t_v is assumed to represent the real value t_v. The spikes that arrive at some node v cause a certain behaviour of P_v that can be represented by a sequence of subsets of v's weights. At each point in time just one member of this sequence needs to be considered to determine whether v fires. We refer to this sequence as the *time course* of the membrane potential of v. For n connections the sequence is known to require not more than $2n$ different subsets as members [4].

The *networks* of spiking neurons that we study here are feedforward networks where the connectivity is defined in terms of a directed acyclic graph. Since we focus on the computation of scalar-valued functions we assume that there is just one node, the *output node*, that has no outgoing connection. Nodes without incoming connections are called *input nodes* and used to represent input vectors to the network. For encoding values as inputs to and outputs from the network binary and analog coding may be used independently. The *depth* of a network is the length of the longest path leading from an input node to the output node. In a *multi-layer network* the nodes are arranged in layers with connections only between subsequent layers. The firing time of a network node v can be described in terms of the firing time of an input node and the delay variables of connections lying on the path leading from the input node to v. We refer to this firing time as the *symbolic output* of node v. Further, we say that this symbolic output is *void* if v does not fire. The symbolic output of a network is defined to be the symbolic output of its output node.

A *dichotomy* of a set $S = \{s^{(1)}, \ldots, s^{(m)}\} \subseteq \mathbb{R}^n$ is a partition of S into disjoint subsets S_0, S_1. Given a set $\mathcal{F}$ of $\{0,1\}$-valued functions we say that $\mathcal{F}$ *shatters* S if $\mathcal{F}$ induces all dichotomies on S. The *Vapnik-Chervonenkis (VC)* *dimension* of $\mathcal{F}$ is defined as the largest number m such that there is a set of m elements that is shattered by $\mathcal{F}$. For a set $\mathcal{F}$ of real-valued functions S is said to be *P-shattered* by $\mathcal{F}$ if there exist real numbers $y^{(1)}, \ldots, y^{(m)}$ such that every dichotomy of $\{(s^{(i)}, y^{(i)}) : i = 1, \ldots, m\}$ is induced by some function of the form $(s, y) \mapsto \text{sign}(f(s) - y)$ for some $f \in \mathcal{F}$. The *pseudo dimension* of $\mathcal{F}$ is the largest number m such that there is a set of m elements that is P-shattered by $\mathcal{F}$. The VC dimension (pseudo dimension) of a network of spiking neurons with binary (analog) coding of the output is is defined to be the VC dimension (pseudo dimension) of the set of functions computed by the network with all possible assignments of values to its parameters.

2. Upper Bounds

In this section we establish upper bounds on the VC and pseudo dimension for networks of spiking neurons as defined in the previous section in terms of depth and number of connections. The networks use analog coding of the inputs and we consider all delays, weights and thresholds as as programmable parameters.

Theorem 1 *A network of spiking neurons with C connections and depth D has VC dimension $O(CD\log(CD))$. This even holds if all delays, weights and thresholds of the network are adjustable.*

Proof. Consider a feedforward network of spiking neurons with one output node, C connections and depth D. Further, let S be an m-element set of input vectors $s^{(1)}, \ldots, s^{(m)}$. The main idea of the proof is to partition the parameter domain of the network, that is, the set of all possible assignment of real values to the network delays, weights and thresholds, into regions such that within each region all parameter values yield the same sequence of symbolic outputs of the network on $s^{(1)}, \ldots, s^{(m)}$. Determining an upper bound on the number of such regions in terms of m, C and D we obtain an upper bound on the number of dichotomies that the network induces on S. Then we use this bound to calculate an upper bound for m when S is shattered by the network.

We proceed by induction on the levels of network nodes. The level of a node v is defined as the length of the longest path leading from an input node to v. Thus, input nodes have level 0 and the level of the output node is equal to the depth of the network.

Let $s \in S$ and consider the set of nodes of a given level λ, say. Assume further that the parameter domain of all nodes of level at most $\lambda - 1$ has already been partitioned into regions such that these nodes do not change their symbolic output on elements from S when the parameters stay within the same region. For each of these regions we partition the parameter domain of the delays of nodes of level at most λ into regions such that within each region the membrane potentials of level λ nodes all have the same time course on the given s. These regions can be bounded by hyperplanes of the form

$$r_{v,i} + d_{v,i} + a \;=\; r_{v,j} + d_{v,j} + b \tag{1}$$

where $r_{v,i}$ and $r_{v,j}$ are symbolic outputs of nodes of level at most $\lambda - 1$ that node v receives through its i-th and j-th connection with corresponding delay parameters $d_{v,i}$ and $d_{v,j}$. Further, a and b are binary values indicating whether the term refers to the starting or ending point of a postsynaptic pulse. Note that $r_{v,i}, r_{v,j}$ are components of s if $\lambda = 1$, and that no hyperplane arises if one of these symbolic outputs is void. If we denote the total number of incoming connections to level λ nodes by $\widehat{C}_\lambda$ then varying over i, j, a and b there are at most $(2\widehat{C}_\lambda)^2$ such hyperplanes that have to be considered for s. Taking all $s \in S$ into account we obtain a number of at most $m(2\widehat{C}_\lambda)^2$ hyperplanes that partition the domain of delay parameters for level λ nodes into regions such that the time courses of these nodes generated on $s^{(1)}, \ldots, s^{(m)}$ do not change when varying their delay parameters within a region.

For each of the regions obtained so far we now define a partition of the parameter domain of weights and thresholds for nodes of level λ. The regions of this partition are chosen in such a way that for each of these nodes the sequence of symbolic outputs on $s^{(1)}, \ldots, s^{(m)}$ does not change when varying its weights and threshold within the same region. This then completes the induction step.

For a particular $s \in S$ these regions arise from hyperplanes of the form

$$\sum_{i \in W_v} w_{v,i} = \theta_v \tag{2}$$

where $w_{v,i}$ and θ_v are weights and threshold of node v and W_v is a subset of the weights occurring in the time course of the membrane potential P_v. We recalled above that at most $2n$ such subsets have to be considered for a node with n inputs and given time course of its membrane potential. Thus at most $2\widehat{C}_\lambda$ such hyperplanes need to be considered for s. Hence, for all $s \in S$ a number of at most $m2\widehat{C}_\lambda$ hyperplanes partition the parameter domain of weights and thresholds of level λ nodes into regions that yield identical sequences of symbolic outputs on S.

It is well known (see, e.g., [4]) that h hyperplanes partition $\mathbb{R}^n$ into at most $2(eh/n)^n$ different regions. (Here, e denotes the base of the natural logarithm.) Applying this to the regions of delay parameters bounded by at most $m(2\widehat{C}_\lambda)^2$ hyperplanes of the form (1), and, for each of these regions, to the regions of weight and threshold parameters bounded by at most $m2\widehat{C}_\lambda$ hyperplanes of the form (2) we obtain an upper bound for the number of regions of the joint parameter domain of delays, weights and thresholds of the nodes of level at most λ by

$$2(em(2\widehat{C}_\lambda)^2/(2C_\lambda + K_\lambda))^{2C_\lambda + K_\lambda} \cdot 2(em2\widehat{C}_\lambda/(2C_\lambda + K_\lambda))^{2C_\lambda + K_\lambda} \tag{3}$$

where K_λ is the number of nodes of level at most λ and C_λ is the total number of incoming connections to these nodes. (In $2C_\lambda + K_\lambda$ we have an upper bound for the number of parameters of these nodes, 2 for each connection and 1 for each node.)

Inductively, we thus obtain an upper bound for the number of regions for the parameter domain of all nodes of level at most l by forming the product of (3) for $\lambda = 1, \ldots, l$. In particular, for the output node of the network, which has level D, this yields the upper bound

$$\prod_{\lambda=1}^{D} 2(em(2\widehat{C}_\lambda)^2/(2C_\lambda + K_\lambda))^{2C_\lambda+K_\lambda} \cdot 2(em2\widehat{C}_\lambda/(2C_\lambda + K_\lambda))^{2C_\lambda+K_\lambda}$$

for the number of regions of the parameter domain of the network such that the sequence of symbolic network outputs on S does not change when varying the parameters within the same region. Using $\widehat{C}_\lambda \leq C_\lambda$, $K_\lambda \leq C_\lambda$ and $C_\lambda \leq C$ we simplify this bound to

$$\prod_{\lambda=1}^{D} 2(2emC)^{3C} \cdot 2(em)^{3C} \quad = \quad 4^D(2e^2m^2C)^{3CD}.$$

Assume now that S is shattered by the network. Then we need to have at least 2^m different functions being computed by the network and hence at least this many different regions of its parameter domain. This implies that

$$m \;\leq\; 2D + 3CD\log(2e^2m^2C)$$

from which we derive $m = O(CD\log(CD))$ by a calculation which is omitted here. This completes the proof of the theorem. $\square$

Along the same lines of reasoning we obtain an upper bound on the pseudo dimension for networks of spiking neurons. The proof is omitted here.

Theorem 2 *A network of spiking neurons with C connections and depth D has pseudo dimension $O(CD\log(CD))$.*

Since the depth of a layered network is equal to the number of its layers we immediately have bounds for multi-layer networks of spiking neurons.

Corollary 3 *A multi-layer network of spiking neurons with C connections and L layers has VC dimension and pseudo dimension $O(CL\log(CL))$.*

3. Lower Bounds

We contrast the results in the previous section with lower bounds that are optimal except for a logarithmic factor. Moreover, these bounds hold even for binary inputs and networks where the delays are the only adjustable parameters.

Theorem 4 *For each $L, C \geq 1$, where $C \geq kL$ for some constant k, there exists a multi-layer network of spiking neurons with L layers, C connections and binary coding of the inputs that has VC dimension $\Omega(CL)$. This even holds if the delays are the only programmable parameters of the network and all weights and thresholds remain fixed.*

Proof. (Sketch) It was shown in [5] that for any $m \in \mathbb{N}$ a module M_m can be constructed that extracts and removes the most significant bit from an m-bit binary number that is given to the module in analog coding. Furthermore, the depth and the number of connections of M_m do not depend on m. We use this module to define a network with $O(L)$ layers and $O(C)$ connections that shatters the set $\{e_i : i = 1, \ldots, C\} \times \{e_j : j = 1, \ldots, L\} \subseteq \{0,1\}^{C+L}$, where e_i denotes the vector with a 1 in the i-th component and 0s elsewhere. □

It was argued in [4] that a function computed by a spiking neuron using binary coding of the inputs can also be computed by such a network using analog coding at the expense of adding one extra input node. Thus the above lower bound passes on to networks of spiking neurons using analog coding of the inputs. Furthermore, the definition of pseudo dimension implies that the same bound also holds for the pseudo dimension of networks of spiking neurons.

The proof of the lower bound uses a network that has fixed weights and thresholds and where the only programmable parameters are C delay variables in the first layer. From this we have the following result.

Corollary 5 *It is impossible to give an upper bound for the VC dimension of a multi-layer network of spiking neurons with fixed weights solely in terms of the number of connections and layers that have programmable delays.*

References

[1] W. Gerstner. Spiking neurons. In W. Maass and C. M. Bishop (eds.), *Pulsed Neural Networks.* MIT Press, Cambridge, Mass., 1999, to app.

[2] W. Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks* **10**, pp. 1659–1671, 1997.

[3] W. Maass. Computing with spiking neurons. In W. Maass and C. M. Bishop (eds.), *Pulsed Neural Networks.* MIT Press, Cambridge, Mass., 1999, to app.

[4] W. Maass and M. Schmitt. On the complexity of learning for a spiking neuron. In *Proc. 10th COLT*, pp. 54–61, ACM, New York, 1997.

[5] W. Maass and M. Schmitt. On the complexity of computing and learning with networks of spiking neurons. In *Proc. 5th Int. Symp. Artificial Intelligence and Mathematics*, http://rutcor.rutgers.edu/~amai/, 1998.

[6] A. F. Murray. Pulse-based computation in VLSI neural networks. In W. Maass and C. M. Bishop (eds.), *Pulsed Neural Networks.* MIT Press, Cambridge, Mass., 1999, to app.

[7] M. Schmitt. Complexity of Boolean computations for a spiking neuron. In L. Niklasson, M. Bodén, and T. Ziemke (eds.), *Proc. 8th ICANN*, vol. 2, pp. 585–590, Springer, London, 1998.

[8] M. Vidyasagar. *A Theory of Learning and Generalization.* Communications and Control Engineering. Springer, London, 1997.