Is there a Support Vector Machine hiding in the dentate gyrus?

John L. Baker
George Mason University, Fairfax VA 22030 USA
Email: jbaker@science.gmu.edu

Abstract

The dentate gyrus has physiological and related behavioral properties suggesting that it implements functions within the hippocampus one of which may be associated with sensory pattern recognition. A top-down dentate gyrus model is defined in terms of an idealized Support Vector Machine pattern recognizer constructed from spiking neurons. The resulting construction offers parallels with dentate gyrus morphology and offers explanation of some of its unique properties, in particular, the mossy fiber pathway and its connection with CA3 pyramidal cells. Derived learning rules suggest properties of the mossy fibers that might be tested experimentally.

Keywords: hippocampus, dentate gyrus, mossy fiber, Support Vector Machine, pattern recognition

1. Introduction

Theories of the hippocampus have assigned a variety of roles to the dentate gyrus. Associative memory models of the hippocampus typically assign the dentate gyrus a role in creating orthogonal input patterns for CA3 so that associative memory saturation is avoided [1,3,7,8]. Lesion studies show that the dentate gyrus is required for spatial learning [6], which suggests a role in sensory pattern recognition.

Explanations as to how the dentate gyrus creates orthogonal patterns for input to CA3 vary, but a common observation is that the number of granule cells in the dentate gyrus is much larger than comparable layer of entorhinal cortex cells. This allows the representation generated in the dentate gyrus to be relatively sparse. Put in other terms, the output of the dentate gyrus has higher dimensionality than its input. This increase in dimensionality is a clue that the formalism of Support Vector Machines may offer some explanation as to the underlying transformation performed in the dentate gyrus. The relative sparseness of connectivity between the dentate gyrus and CA3 is an additional clue suggesting applicability of a Support Vector Machine model.

Support Vector Machines (SVMs) have been described elsewhere [9]. The following description is provided in the interests of a self-contained presentation. Let <*w,x*> be the normal vector dot product of vectors *w* and *x* in Euclidean space. Consider the following class of linear decision functions

$$f(x) = \text{sgn}(< w, x > + b) \in \{-1, 1\} \quad w \in \Re^N, b \in \Re \tag{1}$$

Let the training vectors x_1, x_2, ... be vector-valued inputs and let y_i be the corresponding desired decision function value for the x_i. An optimal decision function is found by solving the optimization problem

$$\text{maximize} \qquad W(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i \alpha_i y_j \alpha_j < x_i, x_j > \tag{2}$$

$$\text{subject to} \qquad \sum_i y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C \quad \text{for some constant } C.$$

The value *C* allows some level of tolerance for when the patterns are not linearly separable. When $\alpha_i = 0$, the associated x_i does not enter into the decision rule. Those x_i for which $\alpha_i > 0$ are termed the support vectors.

SVMs generalize to nonlinear decision functions by embedding the input data in a higher dimensional space termed "Feature Space". Let *F* represent Feature Space and let $\Phi : \Re^N \rightarrow F$ be the map of the input data into Feature Space. Mercer's theorem states that if a function *k* is a positive definite kernel, there is an embedding of *F* in a higher dimensional space with dot products defined as

$$< \Phi(x_i), \Phi(x_j) >= k(x_i, x_j) \tag{3}$$

The function *k* is termed a Mercer kernel. Consider the case where *k* is a function of the dot product $<x_1, x_2>$, which with some abuse of notation we write as $k(x_1, x_2)=k(<x_1, x_2>)$. In this case, a sufficient condition for *k* to form a Mercer kernel is that the Taylor series expansion of *k* as function of $<x_1,x_2>$ have no negative coefficients [10]. This condition is also a necessary condition if the dot product is defined in an infinite dimensional space.

An important consideration for SVMs is that most of the α_i become identically zero and thus do not need to be included in the final decision rule. Support vectors lie on the boundary of decision surface and are sparse with respect to the set of training vectors.

2. Constructing the Support Vector Machine

2.1 Scope of the model

In describing the construction of an SVM from spiking neurons, it is convenient to use terms such as "perforant path", "granule cell" and "CA3 pyramidal" metaphorically. The process is to first derive the model from mathematical constraints and then compare the result with known physiological properties.

2.2 Coding of the perforant path input

For simplicity, perforant path encoding is assumed to be in the form of a population burst. Let x_1, x_2, ... be the various perforant path input spiking patterns where $x_i \in \{0,1\}^N$. For our purposes, each pattern is unique, that is, $i \neq j \Rightarrow x_i \neq x_j$. The index i in this case does not imply temporal sequence.

2.3 Granule cell responses

To provide a response that can be solved analytically, the granule cell is assumed to have a linear *I-V* response [2]. The synaptic response of cell j to input pattern x_i is determined by the dot product $<w_j,x_i>$ where w_j is the vector of synaptic weights. To meet the requirements of the SVM formulation, for a given granule cell, w_j is fixed equal to a unique input pattern selected from among the training vectors.

Cell membrane potential follows the dynamic equation

$$C_m \frac{dV}{dt} = g_{GC}(V_R - V) + X \cdot h_{GC}(t) \tag{4}$$

where: V is the membrane potential, C_m is the membrane capacitance, V_R is the channel reversal potential, g_{GC} is the effective conductance of channels in the granule cell, $X=<w_j,x_i>$, and $h_{GC}(t)$ is the time course of the current from a single synapse.

Let $\lambda=g_{GC}/C_m$ and let V_0 be the initial membrane voltage. We assume without loss of generality that the inputs are presented at time t=0. Equation (4) can then be solved explicitly

$$V(t) = V_R - e^{-\lambda t}(V_R - V_0) + \frac{e^{-\lambda t}}{C_M} \cdot X \cdot \int_0^t e^{\lambda s} h_{GC}(s) ds \tag{5}$$

Let t_f be the time to firing as determined by the time at which $V(t_f)=V_{th}$ where V_{th} is the firing threshold. Because $h(t)$ is assumed to be short, we further assume that $h_{GC}(t)$ is negligible for all t greater than some value less than t_f, or more formally, that there is a $T_H < t_f$ such that $h_{GC}(t)=0$ for all $t>T_H$.

Under this assumption, we can rewrite equation (5) as

$$V(t) = e^{-\lambda t} \cdot (X\beta + V_0 - V_R) + V_R, \quad \beta = \frac{1}{C_m}\int_0^\infty e^{\lambda s} h_{GC}(s)ds, \quad t > T_H \tag{6}$$

From this we then solve for the time of firing as a change in firing time relative to some fixed time T_F where t_f=T_F when X=0.

$$\begin{aligned} t_f &= \frac{1}{\lambda}\ln(\frac{V_R - V_0 - X\beta}{V_R - V_{th}}) = \frac{1}{\lambda}\ln(\frac{V_R - V_0}{V_R - V_{th}}) + \frac{1}{\lambda}\ln(1 - \frac{X\beta}{V_R - V_0}) \\ T_F &= \frac{1}{\lambda}\ln(\frac{V_R - V_0}{V_R - V_{th}}) \\ \delta(X) &= T_F - t_f = -\frac{1}{\lambda}\ln(1 - \frac{X\beta}{V_R - V_0}) \end{aligned} \tag{7}$$

where $\delta(X)$ is the change in firing time and under the condition that $V_R > V_{th} > V_0$.

To check for the dot-product Mercer kernel property, we examine the Taylor series expansion of $\delta(X)$ noting that

$$-\ln(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots \quad -1 < x < 1 \tag{8}$$

The Taylor series of $\delta(X)$ has all positive coefficients and $\delta(\langle w,x\rangle)$ is a Mercer kernel of the input data vector x and the synaptic weight vector w.

Consider the case of a cell with time-varying conductances where a closed form solution for t_f is not available. When conductances are functions of time, a solution for V can be written in the form

$$V(t) = F(t) + X \cdot G(t) \tag{9}$$

If cell conductances vary slowly enough that they can be treated as constant during the interval t_f through T_F, the perturbation of the firing time is similar to that found in equation (7). For X near 0 we have

$$\begin{aligned} \frac{d\delta}{dX} &= G(t_f)\left[\frac{\partial F}{\partial t}(t_f) + X\frac{\partial G}{\partial t}(t_f)\right]^{-1} \\ \delta(X) &= G(T_F)\left(\frac{\partial G}{\partial t}\right)^{-1}\ln\left(1 + X\frac{\partial G}{\partial t}\left(\frac{\partial F}{\partial t}\right)^{-1}\right) \end{aligned} \tag{10}$$

From normal physiology, we expect $G(t)>0$ for excitatory inputs and because the firing threshold is approached from below, we expect $\partial F/\partial t > 0$. A condition for equation (10) resulting in a dot product

Mercer kernel in X would be $\partial G/\partial t < 0$, that is, that the response to synaptic inputs is decreasing when the firing threshold is reached.

2.4 CA3 Pyramidal cell responses

For analytical purposes, a simple model of pyramidal cell response is also used. As before, we assume the existence of one combined ion channel conductance and reversal potential. The membrane potential is modeled by the following dynamic behavior

$$C_m \frac{dV}{dt} = g_{PC}(V_R - V) + \sum_i w_i \cdot h_{MF}(t - t_i) \tag{11}$$

where V is the pyramidal soma membrane potential, C_m is the membrane capacitance, g_{PC} is the net conductance of channels in the pyramidal cell, V_R is the reversal potential, w_1, w_2, w_3, etc. are synaptic weights for the different mossy fiber connections, h_{MF} is the time response of a mossy fiber connection in terms of the current delivered to the soma, and t_1, t_2, t_3, etc. are the times at which action potentials arrive at the various mossy fiber terminals.

We model the mossy fiber terminal as an exponential decay after the initial rise interval

$$h_{MF}(t) = h_N \cdot \exp(-\gamma(t - T_N)) \quad t \geq T_N \tag{12}$$

where: T_N is a time after which the decay is exponential, h_N is a constant chosen to scale the current appropriately, and γ the decay rate. We also have $h_{MF}(t) = 0$ for $t<0$.

Unlike the granule cell, we cannot solve explicitly for the firing time. Instead we examine the voltage at some fixed time sufficiently long after all mossy fiber inputs would have arrived and base our decision function on whether the voltage at that time exceeds a threshold. Let T_B be the time at which we make our measurement, that is, we compare $V(T_B)$ with the threshold voltage V_{th}, and let

$$\begin{aligned} J_i &= \frac{\exp(-\rho T_B)}{C_m} \int_0^{T_B} \exp(\rho s) \cdot h_{MF}(s - t_i) \cdot ds \\ \delta_i(x) &= T_F - t_i, \quad T_F < T_B - T_N, \quad \rho = g_{PC}/C_M \end{aligned} \tag{13}$$

where T_F is the nominal time at which granule cells fire. An explicit solution for $V(t)$ can be found giving a decision function of the form

$$f(x) = \operatorname{sgn}(\sum_i w_i \cdot J_i - b), \quad b = V_{th} - \exp(-\rho T_B) \cdot (V_0 - V_R) - V_R \tag{14}$$

If the function J_i is a Mercer kernel with respect to inputs afferent to the associated granule cell, then equation (15) is in the form of a decision function of a SVM with $w_i = y_i \alpha_i$.

Using the assumed form of h_{MF} and the granule cell response model, we have

$$J_i = \frac{1}{\rho - \gamma}(P\exp(-\gamma\delta_i) - Q\exp(-\rho\delta_i)) + R\cdot\exp(-\rho\delta_i) \tag{15}$$

where

$$\begin{aligned} P &= C_M{}^{-1} h_N \exp(-\gamma T_B + \gamma T_F + \gamma T_N) \\ Q &= C_M{}^{-1} h_N \exp(-\rho T_B + \rho T_F + \rho T_N) \\ R &= C_M{}^{-1} \exp(-\rho T_B + \rho T_F)\int_0^{T_N} \exp(\rho s)\cdot h(s)ds \end{aligned} \tag{16}$$

We can now substitute the granule cell derived values from equation (7). To determine under what conditions the result is a Mercer kernel, we examine the individual exponential forms

$$\begin{aligned} \exp(-\rho\delta_i) &= (1 - aX_i)^{\rho/\lambda} \\ \exp(-\gamma\delta_i) &= (1 - aX_i)^{\gamma/\lambda} \end{aligned} \tag{17}$$

where $a = \beta/(V_R - V_0)$ and V_R and V_0 are taken from the granule cell model in equations (7).

Verification that $1-(1-aX)^r$ forms a dot product Mercer kernel in X for $0<r<1$ and $a>0$ follows from expansion of the Taylor series. Note that because of our assumptions regarding β, we must have $aX<1$. A dot product Mercer kernel does not result when $a>0$ and $r>1$.

There are various conditions under which the J_i could be Mercer kernels of the corresponding X_i:

a) If $\lambda>\gamma>\rho$ and ρ is sufficiently small that the series associated with γ dominates in each Taylor series coefficient. An obvious way to achieve this is to set ρ=0 suppressing leak currents.

b) If $\lambda>\rho>\gamma$ and the series associated with ρ dominates with non-negative coefficients.

c) If $\gamma=\rho$ and ρ is small, then J_i is approximately a linear function of δ_i.

Of these possibilities, γ >ρ appears most consistent with pyramidal cell physiology and mossy fiber responses. Approximate values appear to be γ =0.1/ms [4] and ρ=0.03/ms [11]. If γ/λ =1/2 and the granule cells have normalized synaptic weights, the resulting kernel is equivalent to a Euclidean distance metric.

The signs of mossy fiber synaptic weights have not been considered thus far. While the Lagrangian multipliers for the optimization problem posed in equation (2) cannot be negative, the quantity $y_i\alpha_i$ must necessarily be negative for some *i*. Synaptic weights for support vectors can be made positive, that is, excitatory, through the following algebraic transformation of the decision function.

$$w_i = y_i\alpha_i + C \geq 0, \quad k_I(x) = C \cdot \sum_i k(x, x_i)$$
$$f(x) = \text{sgn}\left(\sum_i w_i k(x, x_i) - k_I(x) - b \right) \tag{18}$$

A physiological correlate to equation (18) would be the requirement for feedforward inhibition of the CA3 pyramidal cell stimulated by granule cell firing. In addition, the time course of the inhibition should approximate that of h_{MF}.

2.5 Learning Rules

Learning corresponds with solving the optimization problem posed in equation (2). Platt [5] has formulated a sequential minimization algorithm (SMO), which makes optimal pair-wise adjustments to the Lagrangian multipliers, or in our terms, mossy fiber synaptic weights. Given the two multiplier values α_1 and α_2, SMO makes adjustments of the form

$$\alpha_2^{new} = \max(L, \min(H, \alpha_2^{old} + \frac{y_2(E_2 - E_1)}{\eta}))$$
$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new}) \tag{19}$$

where

$$y_1 = y_2 \Rightarrow L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C),\ H = \min(C, \alpha_2^{old} + \alpha_1^{old})$$
$$y_1 \neq y_2 \Rightarrow L = \max(0, \alpha_2^{old} - \alpha_1^{old}),\ H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$$
$$\eta = 2k(x_1, x_2) - k(x_1, x_1) - k(x_2, x_2), \quad E_i = f^{old}(x_i) - y_i$$

SMO also provides for adjusting the value of the bias term b in the decision rule whenever one of the α_i is not at an extreme value.

It is doubtful that SMO can be exactly implemented in a neurologically plausible way even in an artificially defined network. As an approximation but at the risk of a solution that is not globally optimal, we can use the SMO rule to set the direction and relative magnitude of weight changes. Heuristically, the initial weight to adjust can be based on the first responding afferent mossy fiber, with adjustments to other afferent weights as needed to maintain a constant sum of mossy fiber synaptic weights. In any case, the adjustment of weights is a form of competitive learning in that increasing one weight involves decreasing others to satisfy the constraint in equation (2).

The maximal response by a granule cell occurs when the input pattern is an exact match for the synaptic weights. If the response values $k(x_i,x_i)$ are normalized to a constant k_{max}, the value of η can be determined from $k(x_1,x_2)$ alone. When the current input pattern is either x_1 or x_2, the value of $k(x_1,x_2)$ is

determined by the difference in firing times of the corresponding mossy fibers. If individual component values of the vector encoded by the perforant path have known statistical properties, for example independence, then an estimator for $k(x_1,x_2)$ can be found based on the difference in firing times. In almost all cases $\eta<0$ and at least the direction of change in synaptic weights can be determined even if the value of η cannot.

Estimating a value for E_i involves knowing the desired outcome of the decision function. This could be accomplished in a number of ways, but in particular through the self-organizing characteristics of CA3 itself. An assumption is that any supervisory stimulus would be reflected in synaptic inputs separate from the mossy fiber connections, for example through the perforant path or recurrent associative connections onto the CA3 pyramidal cell. This results in a learning rule that is dependent not only on the firing of the CA3 cell in question, but also on it's synaptic inputs. Pyramidal cell synaptic input information would need to be conveyed to mossy fiber terminals either via a retrograde messenger or else be extracted from a gradient in the electrical potential at the point of contact between the mossy fiber terminal and the pyramidal cell dendrite.

3. What the SVM construction implies in terms of hippocampus physiology

3.1 Dentate Gyrus

Implementing the SVM construction requires that the dentate gyrus have a set of precisely timed mechanisms. A plausible cycle of events is: 1) a brief inhibition is used to set the granule cell to a known membrane potential, 2) the perforant path synaptic response is integrated in the granule cell, 3) fixed excitatory stimulation is provided to the granule cell such that small input could potentially result in firing, and 4) at time T_F in the cycle, a strong inhibition is applied to prevent late firing. A refinement of step 4 is to use lateral inhibition to suppress further firing once other granule cells have already fired and thus already determined the dominant support vectors. If the inhibition in step 1 is sufficiently punctate, then step 3 can occur earlier or be ongoing, reducing the requirements for precise timing of step 3 stimulation.

Firing of granule cells should occur at a point at which synaptic responses to perforant path input have decreasing effect. Because each granule cell implements a single support vector, learning patterns that require new support vectors would require either the use of previously dormant granule cells or else neurogenesis, in which new granule cells are added. Having a consistent k_{max} would imply a consistent number of synaptic connections between a granule cell and the perforant pathway.

3.2 Mossy fiber pathway

A key requirement of the mossy fiber pathway is that it accurately preserves the relative timing of firing events in the granule cells. This implies that routing of mossy fibers afferent to a given CA3 pyramidal cell follow parallel paths. Outside this parallel routing, granule cells would ideally be arranged in a semi-circular pattern to ensure consistency of axon lengths. To avoid negative synaptic weight values for mossy fiber terminals, mossy fibers must also stimulate inhibitory interneurons that ultimately connect with CA3 pyramidal cells following a path parallel to the mossy fibers themselves.

3.3 Mossy fiber terminals

Mossy fiber terminals at CA3 pyramidal cells would necessarily be complex structures. The efficiency of neurotransmitter release must be very high since relatively few terminals would be active at any one time. In addition, the mossy fiber terminals would have to implement complex learning rules. There is nothing in the construction to differentiate presynaptic and postsynaptic effects, but the existence of a complex presynaptic organelle is consistent with the assumptions made here.

A requirement in implementing a learning rule of the form described is that the mossy fiber terminals exchange information among themselves. This argues for a physically compact arrangement and the use of a variety of extra-cellular messengers. Neurotransmitter spillover could offer one source of exchanged information, but this alone may be insufficient to implement the learning rule needed.

3.4 CA3 Pyramidal cells

In general it is hard to predict which functions would be associated with pyramidal cells and which with the mossy fiber terminals. Retrograde messengers generated in the pyramidal cell would be a reasonable assumption, but no definite messengers are identified. The pyramidal cell is a plausible site for maintaining a consistent number of glutamate receptor sites associated with mossy fiber inputs and thus maintaining a constant sum of total excitatory weight values.

Suppression of pyramidal cell leak currents, at least during the interval during which mossy fiber signals are presented, would enhance the quality of the generated kernel function. Otherwise, the accuracy of the decision function is compromised by leakage during the period over which inputs are integrated.

4. Conclusion

An outline for creating a Support Vector Machine pattern recognizer using spiking neurons has been defined, providing an existence proof that Support Vector Machines are consistent with the constraints of neural physiology. There are implied correspondences between the structures needed for a support vector

recognizer and the dentate gyrus and mossy fiber systems and these structures are not generally found elsewhere in the brain. However, the construction used here is clearly artificial and its success in predicting properties of the hippocampus is a matter for further investigation.

Acknowledgement

I would like to thank Dr. Joel Davis at the Office of Naval Research. This work was supported in part by Office of Naval Research Grant 525424.

References

[1] M.A. Arbib, P. Érdi and J Szentágothai, Neural Organization (MIT Press, Cambridge MA, 1999).

[2] R.A. Fricke and D.A. Prince, Electrophysiology of dentate gyrus granule cells, J. Neuroscience 51 (1984) 195-209.

[3] M.A. Gluck and C.E. Myers, Psychobiological models of hippocampal function in learning and memory, Annu. Rev. Psychol. 48 (1997) 481-514.

[4] D.A. Henze, N.N.Urban and G. Barrionuevo, The multifarious mossy fiber pathway: a review, Neuroscience 98 (2000) 407-427.

[5] J.D. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Schölkopf, C.J.C. Burges and A.J. Smola ed., Advances in Kernel Methods: Support Vector Learning (MIT Press, Cambridge MA, 1999) 185-208.

[6] J-M. Lassalle, T. Bataille and H. Falley, Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task, Neurobiology of Learning and Memory 73 (2000) 243-257.

[7] D. Marr, Simple memory: A theory of the archicortex, Philos. Trans. of the Royal Society of London, Series B 262 (1971) 23-81.

[8] A.D. Redish, Beyond the Cognitive Map (MIT Press, Cambridge MA, 1999).

[9] B. Schölkopf, C.J.C. Burges and A.J. Smola ed., Advances in Kernel Methods: Support Vector Learning (MIT Press, Cambridge MA, 1999).

[10] A.J. Smola, Z.L. Óvarí and R.C. Williamson, Regularization with dot-product kernels, in: T.K. Leen, T.G. Dietterich, and V. Tresp eds., Advanced in Neural Information Processing Systems 13: Proceeding of the 2000 Conference (MIT Press, Cambridge MA, 2001) 308-314.

[11] R.D. Traub, R.K.S. Wong, R. Miles and H. Michelson, A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances, J. Neurophysiol. 66 (1991) 635-650.