

Evolutionary Path to Biological Kernel Machines

Magnus Jändel

Abstract A neural implementation of a support vector machine is described and applied to one-shot trainable pattern recognition. The model is compared to anatomical and dynamical properties of the olfactory system. In the olfactory model, inputs from the olfactory bulb are captured and stabilized in the anterior olfactory cortex and the kernel is computed in the posterior piriform cortex. The anterior piriform cortex contains associative memory populated by support vectors. The associative memory oscillates incessantly between support vector states. Misclassified odours are imprinted as new support vector candidates and the machinery is tuned in sleep. It is demonstrated that there is a plausible evolutionary path from a simple hard-wired pattern recognizer to a full implementation of a biological kernel machine. Simple and individually beneficial modifications are accumulated in each step along this path.

1 One-shot Trainable Pattern Recognition

While an increasing mass of data on brain systems is compiled there is still need for integrative theories of overall function. Learning to recognize new patterns in sensory inputs and to act on such classifications is a key cognitive skill. One single exposure can be sufficient for learning a lesson of high survival interest. Even simple animals such as snails learn food aversion from one exposure [1]. Such one-shot learning is difficult to implement in feed-forward artificial neural networks since vigorous repetition typically is required to build useful skills. Associative memories can, however, learn new patterns instantly [2]. We describe therefore an architecture [3] where significant and surprising experiences are captured to an oscillating associative memory. Memories are tuned and pruned in slow-wave sleep and used for feed-forward pattern recognition in the waking state. It turns out that the system im-

Magnus Jändel
Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden, e-mail: magnus@jaendel.se

plements a support vector machine. This Bio-SVM model is mapped to the anatomy and phenomenology of the olfactory system and it is suggested that many such machines, each tuned to a different context, contribute to odour pattern recognition.

While speculating on intricate mathematical algorithms in living neural systems it is important to remember that complex organisms must have evolved from simpler life forms. A credible model must demonstrate an evolutionary path that starts with a very basic function and gradually builds the complex model in simple steps where each step independently provides add-on survival or reproductive value. A hypothetical evolutionary path to neural support vector machines is therefore presented and discussed in section four. Section two provides the bare-bone mathematical framework of biologically feasible support vector machines. Section three reviews how the model maps to the olfactory system.

2 Kernel Machine Model

Kernel machines or support vector machines (SVM) [4] are efficient pattern recognition algorithms that work by implicitly projecting inputs to a large-dimensional feature space where linear classifiers are applied. The solution is a hyperplane in feature space that separates training example classes with a maximal margin. For simplicity we shall only allow for binary classifications. Consider a set of m training examples (\mathbf{x}_i, y_i) where \mathbf{x}_i is an input vector with binary or real-valued components and $y_i \in \{1, -1\}$ is the correct binary classification of the example. A (zero-bias) SVM classifies a test input vector \mathbf{x} as positive if and only if $f(\mathbf{x}) \geq 0$ where

$$f(\mathbf{x}) = \sum_{i=1}^m y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}). \quad (1)$$

We shall focus on zero-bias v -SVM - a special support vector machine that is uniquely apt for biological implementation [3]. Zero-bias means that there is no constant factor in Eq. (1) as for most support vector machines. The classification function $f(\mathbf{x})$ depends of the training examples, the weights α_i and the non-linear symmetric kernel function K . The weights define the solution to the optimization problem where the dual objective function,

$$W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^m y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j), \quad (2)$$

is maximized subject to,

$$0 \leq \alpha_i \leq \frac{1}{m}, \quad (3)$$

and

$$\sum_{i=1}^m \alpha_i = v. \quad (4)$$

The parameter $0 < \nu < 1$ controls the trade-off between accuracy and generalization. This model is a zero-bias specialization of ν -SVM [5]. The constraint Eq. (4) is applied as suggested by [6]. There are no local optima so the solution to Eqs. (2)-(4) is readily found by gradient ascent in the hyperplane defined by Eq. (4). A simple gradient ascent scheme [3], [7] updates incrementally each weight α_i (subject to Eq. (3)) according to,

$$\Delta\alpha_i \sim \frac{1}{m} \sum_{s=1}^m C_s - C_i, \quad (5)$$

where $\Delta\alpha_i$ is the increment of α_i and C_i is the classification margin of the i :th example,

$$C_i = y_i \sum_{j=1}^m y_j \alpha_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (6)$$

The learning rule (Eq. (5)) drives the weights of easily classified examples to zero. The increment $\Delta\alpha_i$ is always negative if the example is correctly classified with a margin larger than the average margin. Such trivial examples have hence asymptotically vanishing weights $\alpha_i = 0$. The remaining training examples with $\alpha_i > 0$ are called support vectors. Support vectors are borderline events and trivial examples are commonplace events. Only support vectors contribute to classifications. Memory-saving algorithms where trivial examples are discarded from the training set have been shown to be efficient [8].

3 Olfactory Support Vector Machines

Trainable olfactory pattern recognition according to Bio-SVM kernel machine principles is described in [3]. This section reviews the core of the hypothesis. Many different support vector machines classify odours in a wide range of contexts. Each olfactory kernel machine includes memory for support vectors in the anterior piriform cortex (APC), sensory memory for stabilizing inputs in the anterior olfactory cortex (AOC) and classification apparatus in the posterior piriform cortex (PPC). Inputs are provided by the olfactory bulb (OB) and classifications are forwarded to higher-order brain systems (HOBS). HOBS is a place holder for brain systems such as amygdale, the prefrontal cortex, the perirhinal cortex and the entorhinal cortex that are bidirectionally connected to the piriform cortex (PC).

See Fig. 1 for details and the notation that is used in the following subsections. Section 3.1 describes the *Classification process* — how a trained system classifies inputs. The *Surprise learning process* in section 3.2 performs one-shot learning of crucial incidents. Section 3.3 covers the *Importance learning process* where support vector weights are optimized and trivial examples are purged from memory.

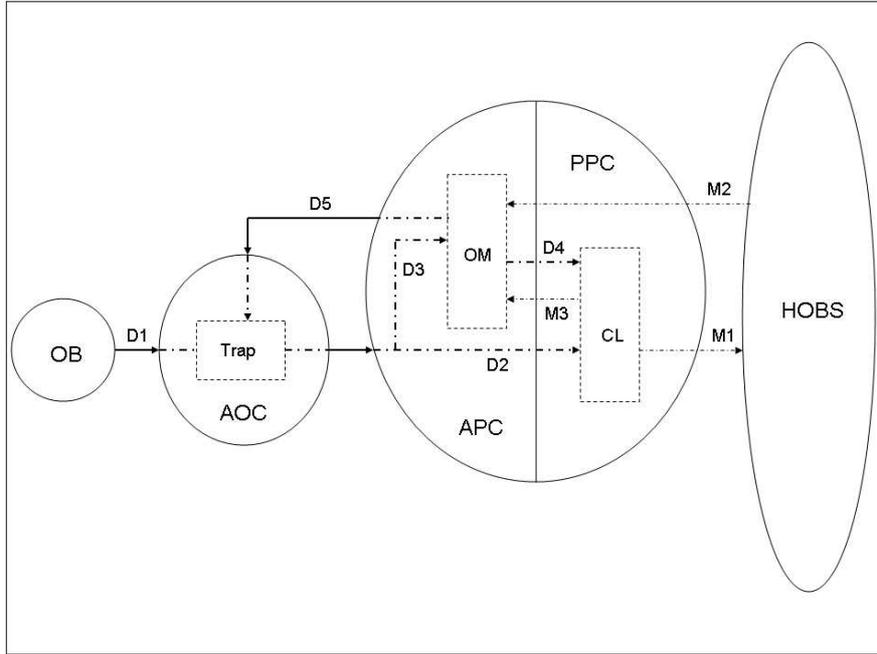


Fig. 1 Outline of an olfactory kernel machine. Solid ovals stand for known brain parts. Higher-order brain systems (HOBS) are management functions in the cortex and the limbic system. OB is the olfactory bulb. AOC is the anterior olfactory cortex. APC and PPC are the anterior and posterior piriform cortex respectively. Dashed boxes indicate hypothetical components of the kernel machine. The Trap is a register for input data in the AOC. OM is oscillating associative memory in the APC and CL is the classification logic in the PPC. Solid lines are known neural projections. Dot-dashed lines are hypothetical connections. Broad connections carrying current or recalled sensory data are D1, D2, D3, D4 and D5. Narrow modulatory projections are M1, M2 and M3. Afferents (D1) carries odour data from the OB to the Trap. Trapped inputs are forwarded to the CL (D2) and to the OM (D3). The OM projects support vectors to the CL (D4) and backwards to the Trap (D5). The CL sends results (M1) to HOBS and learning feedback (M3) to OM. HOBS trigger learning of misclassified examples (M2). The architecture is anatomically feasible but the detailed function is speculative. Note that the figure is highly simplified. Many features that are irrelevant for the present discussion are ignored. See [9] for an overview of the olfactory system.

3.1 The Classification Process

Consider first how a fully trained olfactory support vector machine classifies inputs (D1) from the olfactory bulb. The Trap captures and holds a stable copy \mathbf{x} of the input for the duration of a sniff cycle of 125 – 250 ms. The OM is an associative memory for support vectors. It oscillates rapidly between support vector states with a frequency much faster than the sniff cycle. See [10], [11] and [12] for simulations showing the feasibility of such oscillating memories and [3] for an in-depth discussion of the OM.

The OM displays a memory state for a short time before it oscillates to the next state. The endurance time T_i is the average duration of memory state \mathbf{x}_i in the perpetual oscillation of the OM. The endurance time is the physical parameter that encodes the support vector weight of the memory state. In the following we use α_i as a shorthand for a dimensionless parameter that is proportional to T_i and plays the part of the SVM weight of the training example that is engraved as memory state \mathbf{x}_i .

The SVM kernel function $K(\mathbf{x}_i, \mathbf{x})$ is computed in the CL where projections carrying support vectors \mathbf{x}_i (D4) join afferents (D2) conveying the input vector \mathbf{x} . The classification function Eq. (1) is computed by temporal summation,

$$f(\mathbf{x}) \sim \int_{t_0}^{t_0+T_{trap}} y_{i(t)} K(\mathbf{x}_{i(t)}, \mathbf{x}) dt, \quad (7)$$

where t_0 is the starting time of the integration, T_{trap} is the holding time of the sensory memory and $i(t)$ is the index of the prevailing OM memory pattern at time t . As usual, y_i is the valence of the memory pattern \mathbf{x}_i . The resulting classification is transmitted to HOBS (M1). Note that neural temporal summation produces an approximation of Eq. (1) where the non-linear summation of physical neurons and the stochastic nature of the presentation of the support vectors contribute to pattern recognition errors. Further details of the classification process will be discussed in section 4.

3.2 Surprise Learning

In this paper, we define a surprise as a stimulus that causes a neural classifier to make an error. New stimuli may be correctly classified but a surprise is by definition misclassified. Misclassifications cause strong emotional responses with positive or negative valence and trigger a surge of neuromodulators (M2) causing the OM to engrave the misclassified pattern as a new support vector candidate. The Trap holds a stable copy of the surprising input that projects (D3) to the OM. The emotional valence of the surprise provides the label y_i of the new memory pattern \mathbf{x}_i . Note that the new training example is learned from one single exposure. Mechanisms for such one-shot learning in artificial associative memories are described by Hopfield [2]. The SVM weights α_i are sub-optimal following the addition of a new training example so the animal may not classify scents correctly immediately after misclassification events.

3.3 Importance Learning

From ancient times it has been surmised that memory is trimmed and consolidated in sleep [13]. We suggest a specific application of this idea - support vector weights are optimized and trivial examples are pruned from memory while the animal sleeps. As external inputs are suppressed in sleep, the Trap locks on inputs (D5) from the OM. Real-world data are replaced with support vectors. The OM keeps oscillating incessantly in the sleeping brain so that support vectors are presented stochastically. The Trap holds each such training example (\mathbf{x}_j, y_j) for the duration of a sniff cycle and will then capture the next support vector that is presented by the OM. The OM oscillates much faster than the sniff cycle. The probability of trapping any given example i is hence proportional to the corresponding endurance time T_i . The CL computes the kernel $K(\mathbf{x}_i, \mathbf{x}_j)$. Note that (\mathbf{x}_j, y_j) is the example that is trapped and (\mathbf{x}_i, y_i) is the example that currently is offered by the OM. A feedback signal $B_{ij} = y_j K(\mathbf{x}_i, \mathbf{x}_j)$ is projected (M3) from the CL to the OM. Note that the kernel computation in the CL thus has dual use.

Once during each OM oscillation the learning rules,

$$\Delta T_i \sim -y_i B_{ij} \quad \text{and} \quad \forall s : \Delta T_s \sim \frac{1}{m} y_i B_{ij}, \quad (8)$$

are applied. In Eq. (8), ΔT_s is the increment of the endurance time T_s . The current memory pattern i is hence depressed in proportion to $y_i B_{ij}$ and all memory patterns are potentiated in proportion to $\frac{1}{m} y_i B_{ij}$. The sum of endurance times is conserved. Averaging Eq. (8), for any given memory pattern i , over the probability distribution of the trapped examples j gives the effective learning rules,

$$\widetilde{\Delta T}_i \sim -C_i \quad \text{and} \quad \forall s : \widetilde{\Delta T}_s \sim \frac{1}{m} C_i, \quad (9)$$

where $\widetilde{\Delta T}_s$ is the average increment of T_s and C_i is given by Eq. (6). Note that the support vector weight α_s is proportional to T_s . The OM implements hence zero-bias v-SVM gradient ascent according to Eq. (5). This means that the biological support vector machine eventually acquires optimal weights. Trivial examples are erased from the OM as the corresponding weights fall to zero.

4 Evolutionary Path

This section describes a hypothetical evolutionary path from primitive pattern recognition to a full implementation of kernel machines in low-level perception. The path consists of a sequence of simple modifications where each step brings some advantage to the life form. As an ongoing example we shall consider an organism living in a world with many different food stuffs and many different toxic substances. Using

odours for distinguishing food from poison is crucial. As organisms evolve they will be equipped with increasingly sophisticated chemical pattern recognition systems.

The primordial pattern recognition system consists of a sensor system SS and a pattern recognizer PR (Fig. 2a). The sensor system includes receptor cells and back-end layers for stabilizing and filtering the external input. A prototypical sensor system is the primary receptor cells combined with the glomerular layer of the olfactory bulb. The output of the pattern recognizer is a function $f(\mathbf{x}''(t))$ where \mathbf{x}'' is the sensory input vector and t is time. A positive value of f could e.g. mean safe to eat while a negative value indicates not safe to eat.

Turbulence in the odour carrying medium causes discontinuous and highly variable exposure of odour signals at the chemoreceptor neurons [14]. Adding sensory memory SM enables more sophisticated analysis (Fig. 2b). SM captures a snapshot \mathbf{x} of the sensor signal and holds it stable for a time T_{trap} until the next snapshot is trapped. More time is now available for computing a complex classification function $f(\mathbf{x})$ of a significant input \mathbf{x} . Sensory memory expands the range of features and phenomena that the system can recognize.

The organism can learn to recognize new scents by adapting neural networks in the PR thus modifying $f(\mathbf{x})$. Learning new pattern recognition skills by tweaking $f(\mathbf{x})$ means, however, that new memories overwrites old unless all relevant training examples are repeated continuously.

One-shot learning is an essential skill in a world where organisms can not afford to repeat mistakes. As a starting point for evolving one-shot learning, we assume that associative memory (AM) is available in the brain (Fig. 2c). This facility has developed for some other purpose and is initially disconnected from the pattern recognition system.

The next evolutionary step is to connect SM to the AM (Fig. 2d). Frightening, painful, pleasurable or otherwise emotional events cause a burst of neuromodulators that imprints the present sensory input \mathbf{x} as a new memory pattern in the AM. Hopfield demonstrated the feasibility of such one-shot learning in a model of associative memory [2]. The emotional valence y of the input is a part of the memory trace. In the food search example $y = 1$ indicates food and $y = -1$ means poison. Significance events are hence represented in persistent memory by the associated input pattern and the emotional valence. Sensory memory is essential for temporary saving the input that caused the surprise.

As the organism explores its environment, the input \mathbf{x} falls within the basin of attraction of memory pattern \mathbf{x}' and cues the AM to settle into the state (\mathbf{x}', y') . The emotional valence y' of the triggered memory is produced (see Fig. 2d). Remembering the emotional valence of a training example that is similar to the presently encountered substance helps to select food and avoid poisons. The system can use y' directly to drive actions or more likely fuse it with other evidence in high-level decision modules.

The system of Fig. 2d would work rather well in a world where all substances are known and have unambiguous sensory signatures. In a less clear-cut environment, food scents that are only remotely similar to a known poison could fall into the

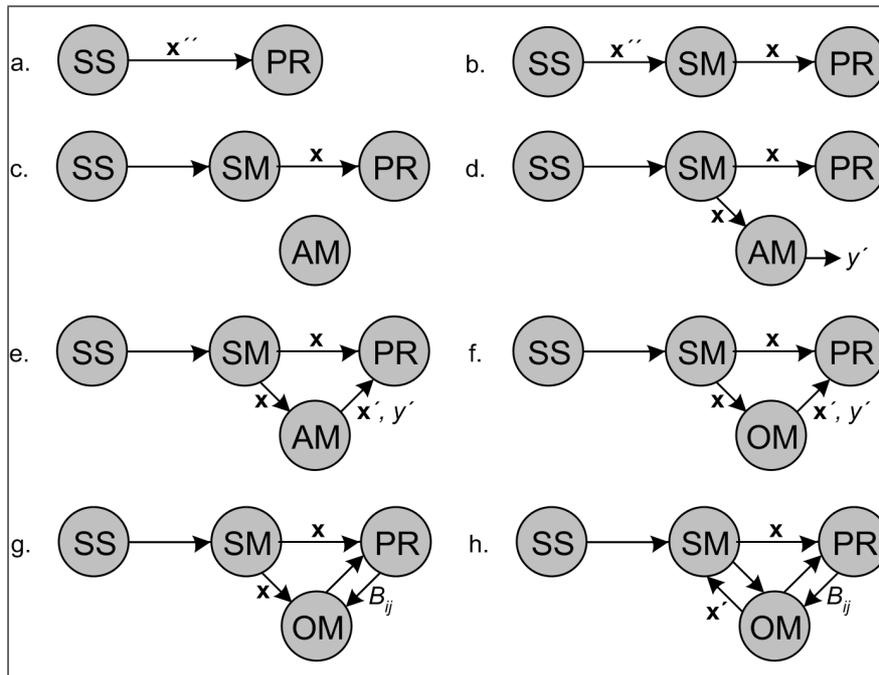


Fig. 2 Speculative evolutionary path leading to a biological kernel machine according to the Bio-SVM concept. The key adaptation is indicated for each step.

- a) Base-line pattern recognition system consisting of a sensor system (SS) and a pattern recognizer (PR). The signal from SS to PR is the sensory vector x'' .
- b) The system is extended with sensory memory (SM) providing a stable duplicate x of the sensory vector.
- c) Associative memory (AM) is available in the brain.
- d) Surprising signals from SM are stored in AM. The emotional valence y is recalled for sufficiently similar inputs.
- e) The PR modulates the recalled valence with a similarity measure comparing x with the stored pattern x' .
- f) Oscillating memory (OM) and temporal summation in the PR enable pattern recognition based on a weighted average over many training examples.
- g) Learning feedback from the PR to the OM tunes memory weights in real-world experiments.
- h) Feedback from the OM to the SM enables virtual experiments in sleep thus completing a biological support vector machine.

basin of attraction of the corresponding memory state and thus trigger unwarranted avoidance behaviour.

A successful mutation could build a connection from the associative memory to the pattern recognition module (Fig. 2e). The input \mathbf{x} falls within the basin of attraction of some memory pattern \mathbf{x}' and causes the AM to settle in the state \mathbf{x}' . The PR receives \mathbf{x} , \mathbf{x}' and y' . The classification function would now be of the form $f(\mathbf{x}) = y'K(\mathbf{x}', \mathbf{x})$ where the function K measures the similarity of \mathbf{x}' and \mathbf{x} . The PR outputs the recalled valence y tempered by a measure of similarity between the present sensory signal and the recalled example. An activation threshold could ensure that only sufficient similar \mathbf{x} and \mathbf{x}' trigger actions suggested by y' . The resulting behaviour would be more appropriate e.g. with feeding triggered only by substances that are quite similar to known foods. A disadvantage of this system is that the organism gets little guidance if the selected pattern \mathbf{x}' is too dissimilar to \mathbf{x} since the system compares the input with just one of the training examples.

The next evolutionary step is to compare the sensory input to many stored patterns. To achieve this, the associative memory transforms so that it will not settle into a stable attractor but rather perpetually oscillate between memory states. The associative memory becomes an oscillating memory OM (Fig. 2f). The wide-ranging phenomenon of chaotic itinerancy ([15], [16], [17] see [18] for a review) lends credibility to the existence of such oscillating memories in brains and shows that a minor change in the dynamics of biological associative memory can cause a transition to the oscillating phase.

The pattern recognizer employs temporal summation to compute

$$f(\mathbf{x}) \sim \int_{t_0}^{t_0+T_{trap}} y_{i(t)} K(\mathbf{x}_{i(t)}, \mathbf{x}) dt, \quad (10)$$

where t_0 is the starting time of the integration, T_{trap} is the holding time of the sensory memory, $i(t)$ is the index of the present memory pattern of the OM and y_i is the valence of the memory pattern \mathbf{x}_i . Temporal summation is a naturally occurring property of neurons [19] and may already be available in the PR although it served no computational function in preceding systems. No change to the PR may hence be required in the transition to the system of Fig. 2f.

If T_{trap} is much larger than the OM oscillation time, Eq. (10) averages to,

$$f(\mathbf{x}) \approx c \sum_{i=1}^m y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}), \quad (11)$$

where c is a positive constant and α_i are weights proportional to the endurance times T_i of the corresponding memory patterns. The endurance time could depend on the emotional intensity of the event that imprinted the corresponding memory trace.

Our organism can now perform pattern recognition based on weighted averages of similarity measures for many stored memories. It has in fact implemented the classification process of a support vector machine as described in sections 2 and 3. Eq. (11) is identical to Eq. (1) provided that K is understood as the kernel function of the support vector machine, α_i are the SVM weights and the signum function is

applied for binary classifications. The Surprise learning process of section 3 is also identical to one-shot learning as described in this section. Further evolution could explore that the same oscillating memory can serve multiple pattern recognition units, each tailored for a different purpose.

Temporal integration according to Eq. (10) will, however, converge within reasonable time only for a limited number of memory patterns. The capacity of the oscillating memory is also finite. To handle rich and variable environments the organism needs means for trimming the content of the OM to a small and dynamically updated population of vital training examples.

The next evolutionary invention is to carry a feedback signal $B_{ij} = y_j K(\mathbf{x}_i, \mathbf{x}_j)$ from the PR to the OM (Fig. 2g). Note that (\mathbf{x}_i, y_i) indicates the state of the OM while \mathbf{x}_j is the sensory vector held by the SM. The feedback signal B_{ij} includes the valence y_j as evaluated by higher-order brain systems in interaction with the world. Tasting the substance with scent \mathbf{x}_j gives e.g. the classification y_j (edible or toxic). The OM uses B_{ij} to regulate endurance times according to Eq. 8.

Consider a world with \hat{m} substances that the organism encounters with probability \hat{p}_j . Each substance is either edible or toxic. Averaging B_{ij} for a given OM state i over world states j and multiplying with the valence y_i of the OM state gives $\hat{C}_i = y_i \sum_{j=1}^{\hat{m}} y_j \hat{p}_j K(\mathbf{x}_i, \mathbf{x}_j)$. Note that \hat{C}_i is the classification margin of \mathbf{x}_i for the classifier $\hat{f}(\mathbf{x}) = \sum_{j=1}^{\hat{m}} y_j \hat{p}_j K(\mathbf{x}_j, \mathbf{x})$ that averages over valences of real-world substances weighted with real-world probability \hat{p}_j and the similarity measure $K(\mathbf{x}_j, \mathbf{x})$. We also define \hat{C} averaged over all m memory patterns in the OM, $\hat{C}_{OM} = \frac{1}{m} \sum_{s=1}^m \hat{C}_s$. Applying the learning rules in Eq. 8 means that the endurance time T_i of all memory states with $\hat{C}_i > \hat{C}_{OM}$ are driven to zero. Such states are hence pruned from the OM. The endurance time of states with $\hat{C}_i < \hat{C}_{OM}$ are pushed to the maximum value $T_i = T_{max}$.

The effect of adding the feedback $B_{ij} = y_j K(\mathbf{x}_i, \mathbf{x}_j)$ (Fig. 2g) and applying the OM learning rules is that training examples that are correctly classified with a good margin are purged from the OM. The system retains training examples with a narrow margin that are hard to classify correctly. Such examples mark the borderline between categories and provide hence useful information for classification purposes. Dropping the high-margin trivial examples subtracts little from pattern recognition performance but makes the system much faster and reduces the need for memory capacity. The selection of training examples and the associated endurance times (weights) is, however, not optimal. The classifier $\hat{f}(\mathbf{x})$ is different from $f(\mathbf{x})$ so the learning process will drive the OM population and weights to a suboptimal state from a support vector machine point of view.

Evolution completes the implementation of a biological support vector machine by adding a backward projection carrying memory patterns from the OM to SM (Fig. 2h). Input from the sensors SS dominates, however, in the waking state. As sensors are turned off in sleep, the SM will trap the otherwise suppressed input from the backward projection. Randomly selected training examples masquerade for actual sensory data. The optimization of the SVM weights is performed using the same learning rules as in the system of Fig. 2g but replacing real-world inputs with recalled training examples as described in section 3.3. This final step provides

two major advantages. Firstly, pattern recognition performance is improved since the selection of OM training examples and their weights is optimal. Secondly, the optimization process proceeds swiftly by virtual experimentation in sleep rather than by slow and dangerous trial-and-error in the real world.

4.1 Conclusions and Discussion

It should be understood that early rungs of the evolutionary ladder could coexist with modern structures. The direct connection from the olfactory bulb to the piriform cortex might e.g. be a part of a legacy discrimination system (at the level of Fig. 2a) that is hard-wired for detecting scents of high survival significance [HAB]. Later evolutionary steps employ, according to section 3, the rewired route through the anterior olfactory cortex. The lobster olfactory system includes also labelled lines where dedicated subsystems handle specific odorants of particular survival value [20].

The locus for short-term sensory memory could also be in the olfactory bulb. Periodic signalling of the memory state to a secondary sensory memory in the anterior olfactory cortex would be consistent with the present model. The procerebral lobe, that has a similar function as the olfactory bulb in invertebrate species, seems to be the site of odour sensory memory [21]. It is also conceivable that several types of sensory memory with different time scales operate in different contexts.

Section 3 reviews a new model for olfactory pattern recognition. Note, however, that there is a wealth of computational approaches to olfaction (see [22] for a review). Odour recognition models based on cortical dynamics include [23]. Associative memory in the piriform cortex is described by [24]. Models with central information processing in the olfactory bulb include [25], [26].

The Bio-SVM model of trainable pattern recognition matches the architecture of both the thalamic [7] and the olfactory system [3]. It appears that there is an evolutionary path from simple hard-wired pattern recognition to the full Bio-SVM architecture. Each step along the path includes one single modification in components or connections that provides some crucial advantage in pattern recognition performance. If evolution actually has travelled along this path is an open issue.

Acknowledgements This work was supported by the Swedish Foundation for Strategic Research. Enlightening discussions with Hans Liljenström are gratefully acknowledged.

References

1. Teyke, T.: Food-attraction conditioning in the snail, *Helix Pomatia*. *J. Comp. Physiol. A* **177**, 409–414 (1995)
2. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Sci. USA* **79**, 2554–2558 (1982)

3. Jändel, M.: A neural support vector machine. *Neural Networks*, in press <http://dx.doi.org/10.1016/j.neunet.2010.01.002> (2010)
4. Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines and Other Kernel-Based Methods*. Cambridge University Press, Cambridge (2000)
5. Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. *Neural Computation* **12**, 1207–1245 (2000)
6. Chang, C-C., Lin, C-J.: Training ν -support vector classifiers: theory and algorithms. *Neural Computation* **13**, 2119–2147 (2001)
7. Jändel, M.: Thalamic bursts mediate pattern recognition. *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering* 562–565 (2009)
8. Schölkopf, B., Smola, A.J.: *Learning with Kernels*. MIT Press, Cambridge, MA (2002)
9. Haberly, L.B.: Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses* **26**, 551–576 (2001)
10. Pantic, L., Torres, J.J., Kappen, H.J., Gielen, S.: Associative memory with dynamic synapses. *Neural Computation* **14**, 2903–2923 (2002)
11. Horn, D., Usher, M.: Neural networks with dynamical thresholds. *Physical Review A* **40(2)**, 1036–1044 (1989)
12. Liljenström, H.: Neural stability and flexibility: a computational approach. *Int. J. Neuropsychopharmacol.* **28**, 64–73 (2003)
13. Quintilianus, M.F.: *Institutio Oratoria*, Book XI (English translation in *The Orators Education*, Vol. **5**, Books 11–12, Loeb classical library) (95)
14. Koehl, M.A.R., Koseff, J.R., Grimaldi, J.P., McCay, M.G., Cooper, T., Wiley, M.B., Moore, P.A.: Lobster sniffing: antennule design and hydrodynamic filtering of information in an odor plume. *Science* **294**, 1948–1951 (2001)
15. Ikeda, K., Matsumoto, K., Otsuka, K.: Maxwell-Bloch turbulence. *Prog. Theor. Phys. Suppl.* **99**, 295–324 (1989)
16. Kaneko, K.: Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements. *Physica D* **41**, 137–172 (1990)
17. Tsuda, I.: Dynamic link of memory: chaotic memory map in nonequilibrium neural networks. *Neural Networks* **5**, 313–326 (1992)
18. Kaneko, K., Tsuda, I.: Chaotic itinerancy. *Chaos* **13**, 926–936 (2003)
19. Johnston, D., Wu, S.M-S.: *Foundations of Cellular Neurophysiology*. MIT Press, Cambridge MA (1995)
20. Derby, C.D.: Learning from spiny lobsters about chemosensory coding of mixtures. *Physiol. Behav.* **69**, 203–209 (2000)
21. Gelperin, A.: Oscillatory dynamics and information processing in olfactory systems. *J. Exp. Biology* **202**, 1855–1864 (1999)
22. Cleland, T.A., Linster, C.: Computation in the olfactory system. *Chem. Senses* **30**, 801–813 (2005)
23. Liljenström, H.: Modeling the dynamics of olfactory cortex using simplified network units and realistic architecture. *Int. J. Neural Syst.* **2**, 1–15 (1991)
24. Li, Z., Hertz, J.: Odour recognition and segmentation by a model olfactory bulb and cortex. *Network: Comput. Neural Syst.* **11**, 83–102 (2000)
25. Freeman W.J.: *Mass action in the Nervous System*. Academic Press, New York (1975)
26. Skarda, C.A., Freeman W.J.: How brains make chaos to make sense of the world. *Behav. Brain Sci.* **10**, 161–195 (1987)