Linear readout of object manifolds

SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky

Linear Readout of Object Manifolds

SueYeon Chung,[1, 2] Daniel D. Lee,[3] and Haim Sompolinsky[2, 4, 5, *]

[1]*Program in Applied Physics, School of Engineering and Applied Sciences,*
Harvard University, Cambridge, MA 02138, USA

[2]*Center for Brain Science, Harvard University, Cambridge, MA 02138, USA*

[3]*Department of Electrical and Systems Engineering,*
University of Pennsylvania, Philadelphia, PA 19104, USA

[4]*Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel*

[5]*Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem 91904, Israel*

Objects are represented in sensory systems by continuous manifolds due to sensitivity of neuronal responses to changes in physical features such as location, orientation, and intensity. What makes certain sensory representations better suited for invariant decoding of objects by downstream networks? We present a theory that characterizes the ability of a linear readout network, the perceptron, to classify objects from variable neural responses. We show how the readout perceptron capacity depends on the dimensionality, size, and shape of the object manifolds in its input neural representation.

High-level perception in the brain involves classifying or identifying objects which are represented by continuous manifolds of neuronal states in all stages of sensory hierarchies [1–7] Each state in an object manifold corresponds to the vector of firing rates of responses to a particular variant of physical attributes which do not change object's identity, e.g., intensity, location, scale, and orientation. It has been hypothesized that object identity can be decoded from high level representations, but not from low level ones, by simple downstream readout networks [1, 2, 6, 8–12]. A particularly simple decoder is the perceptron, which performs classification by thresholding a linear weighted sum of its input activities [13, 14]. However, it is unclear what makes certain representations well suited for invariant decoding by simple readouts such as perceptrons. Similar questions apply to the hierarchy of artificial deep neural networks for object recognition [10, 15–18]. Thus, a complete theory of perception requires characterizing the ability of linear readout networks to classify objects from variable neural responses in their upstream layer.

A theoretical understanding of the perceptron was pioneered by Elizabeth Gardner who formulated it as a statistical mechanics problem and analyzed it using replica theory [19–28]. In this work, we generalize the statistical mechanical analysis and establish a theory of linear classification of manifolds synthesizing statistical and geometric properties of high dimensional signals. We apply the theory to simple classes of manifolds and show how changes in the dimensionality, size, and shape of the object manifolds affect their readout by downstream perceptrons.

Line segments: One-dimensional object manifolds arise naturally from variation of stimulus intensity, such as visual contrast, which leads to approximate linear modulation of the neuronal responses of each object. We model these manifolds as line segments and consider classifying P such segments in N dimensions, expressed as $\{\mathbf{x}^\mu + Rs\mathbf{u}^\mu\}$, $-1 \leq s \leq 1$, $\mu = 1, ..., P$. The N-dimensional vectors $\mathbf{x}^\mu \in \mathcal{R}^N$ and $\mathbf{u}^\mu \in \mathcal{R}^N$ denote respectively, the *centers* and *directions* of the μ-th segment, and the scalar s parameterizes the continuum of points along the segment. The parameter R measures the extent of the segments relative to the distance between the centers (Fig. 1).

We seek to partition the different line segments into two classes defined by binary labels $y^\mu = \pm 1$. To classify the segments, a weight vector $\mathbf{w} \in \mathcal{R}^N$ must obey $y^\mu \mathbf{w} \cdot (\mathbf{x}^\mu + Rs\mathbf{u}^\mu) \geq \kappa$ for all μ and s. The parameter $\kappa \geq 0$ is known as the margin; in general, a larger κ indicates that the perceptron solution will be more robust to noise and display better generalization properties [29]. Hence, we are interested in maximum margin solutions, i.e., weight vectors $\mathbf{w}$ that yield the maximum possible value for κ. Since line segments are convex, only the endpoints of each line segment need to be checked, namely min $h_0^\mu \pm Rh^\mu = h_0^\mu - R|h^\mu| \geq \kappa$ where $h_0^\mu = ||\mathbf{w}||^{-1}y^\mu \mathbf{w} \cdot \mathbf{x}^\mu$ are the fields induced by the centers and $h^\mu = ||\mathbf{w}||^{-1}y^\mu \mathbf{w} \cdot \mathbf{u}^\mu$ are the fields induced by the line directions.

Replica theory: The existence of a weight vector $\mathbf{w}$ that can successfully classify the line segments depends upon the statistics of the segments. We consider random line segments where the components of $\mathbf{x}^\mu$ and $\mathbf{u}^\mu$ are i.i.d. Gaussians with zero mean and unit variance, and random binary labels y^μ. We study the thermodynamic limit where the dimensionality $N \to \infty$ and number of segments $P \to \infty$ with finite $\alpha = P/N$ and R. Following Gardner [19] we compute the average of $\log V$ where V
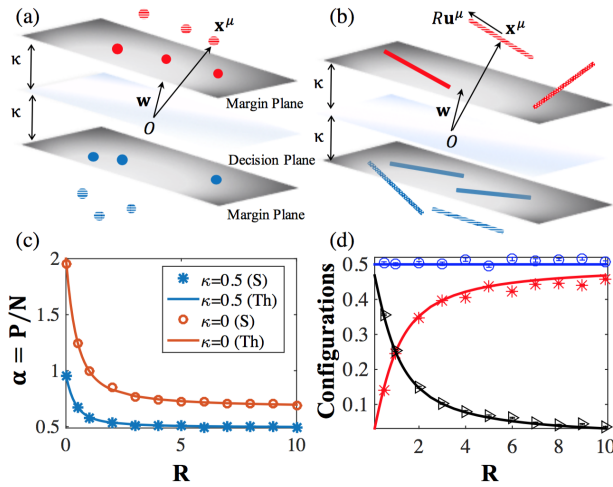
FIG. 1: (a) Linear classification of points. (solid) points on the margin, (striped) internal points. (b) Linear classification of line segments. (solid) lines embedded in the margin, (dotted) lines touching the margin, (striped) interior lines. (c) Capacity $\alpha = P/N$ of a network $N = 200$ as a function of R with margins $\kappa = 0$ (red) and $\kappa = 0.5$ (blue). Theoretical predictions (lines) and numerical simulation (markers, see SM for details) are shown. (d) Fraction of different line configurations at capacity with $\kappa = 0$. (red) lines in the margin, (blue) lines touching the margin, (black) internal lines.

is the volume of the space of perceptron solutions:

$$V = \int_{\|\mathbf{w}\|^2 = N} d^N \mathbf{w} \prod_{\mu=1}^{P} \Theta \left(h_0^\mu - R \left| h^\mu \right| - \kappa \right). \quad (1)$$

$\Theta(x)$ is the Heaviside step function. According to replica theory, the fields are described as sums of random Gaussian fields $h_0^\mu = t_0^\mu + z_0^\mu$ and $h^\mu = t^\mu + z^\mu$ where t_0 and t are quenched components arising from fluctuations in the input vectors $\mathbf{x}^\mu$ and $\mathbf{u}^\mu$ respectively, and the z_0, z fields represent the variability in h_0^μ and h^μ resulting from different solutions of $\mathbf{w}$. These fields must obey the constraint $z_0 + t_0 - R|z + t| \geq \kappa$. The capacity function $\alpha_1(\kappa, R)$ (the subscript 1 denotes the dimensionality of the manifolds) describes for which P/N ratio the perceptron solution volume shrinks to a unique weight vector. The reciprocal of the capacity is given by the replica symmetric calculation (details provided in supplementary materials, SM):

$$\alpha_1^{-1}(\kappa, R) = \left\langle \min_{z_0 + t_0 - R|z + t| \geq \kappa} \frac{1}{2} \left[z_0^2 + z^2 \right] \right\rangle_{t_0, t} \quad (2)$$

where the average is over the Gaussian statistics of t_0 and t. To compute Eq. (2), three regimes need to be considered. First, when t_0 is large enough so that $t_0 > \kappa + R|t|$, the minimum occurs at $z_0 = z = 0$ which does not contribute to the capacity. In this regime, $h_0^\mu > \kappa$ and $h^\mu > 0$ implying that neither of the two segment endpoints reach the margin. In the other extreme, when $t_0 < \kappa - R^{-1}|t|$,

the minimum is given by $z_0 = \kappa - t_0$ and $z = -|t|$, i.e. $h_0^\mu = \kappa$ and $h^\mu = 0$ indicating that both endpoints of the line segment lie on the margin planes. In the intermediate regime where $\kappa - R^{-1}|t| < t_0 < \kappa + R|t|$, $z_0 = \kappa - t_0$ but $z > -|t|$, i.e. $h_0^\mu = \kappa$ but $h^\mu > 0$, corresponding to only one of the line segment endpoints touching the margin. In this regime, the solution is given by minimizing the function $(R|z + t| + \kappa - t_0)^2 + z^2$ with respect to z. Combining these contributions, we can write the perceptron capacity of line segments:

$$\alpha_1^{-1}(\kappa, R) = \int_{-\infty}^{\infty} Dt \int_{\kappa - R^{-1}|t|}^{\kappa + R|t|} Dt_0 \frac{\left(R|t| + \kappa - t_0 \right)^2}{R^2 + 1}$$

$$+ \int_{-\infty}^{\infty} Dt \int_{-\infty}^{\kappa - R^{-1}|t|} Dt_0 \left[(\kappa - t_0)^2 + t^2 \right] \quad (3)$$

with integrations over the Gaussian measure, $Dx \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$. It is instructive to consider special limits. When $R \to 0$, Eq. (3) reduces to $\alpha_1(\kappa, 0) = \alpha_0(\kappa)$ where $\alpha_0(\kappa)$ is Gardner's original capacity result for perceptrons classifying P points (the subscript 0 stands for zero-dimensional manifolds) with margin κ 1-(a). Interestingly, when $R = 1$, then $\alpha_1(\kappa, 1) = \frac{1}{2}\alpha_0(\kappa/\sqrt{2})$. This is because when $R = 1$ there are no statistical correlations between the line segment endpoints and the problem becomes equivalent to classifying $2P$ random points with average norm $\sqrt{2N}$.

Finally, when $R \to \infty$, the capacity is further reduced: $\alpha_1^{-1}(\kappa, \infty) = \alpha_0^{-1}(\kappa) + 1$. This is because when R is large, the segments become unbounded lines. In this case, the only solution is for $\mathbf{w}$ to be orthogonal to all P line directions. The problem is then equivalent to classifying P center points in the $N - P$ null space of the line directions, so that at capacity $P = \alpha_0(\kappa)(N - P)$.

We see this most simply at zero margin, $\kappa = 0$. In this case, Eq. (3) reduces to a simple analytic expression for the capacity: $\alpha_1^{-1}(0, R) = \frac{1}{2} + \frac{2}{\pi} \arctan R$ (SM). The capacity is seen to decrease from $\alpha_1(0, R = 0) = 2$ to $\alpha_1(0, R = 1) = 1$ and $\alpha_1(0, R = \infty) = \frac{2}{3}$ for unbounded lines, and can be related to a perceptron model with varying correlations between pairs of points [28]. We have also calculated analytically the distribution of the center and direction fields h_0^μ and h^μ [27]. The distribution consists of three contributions, corresponding to the regimes that determine the capacity. One component corresponds to line segments fully embedded in these planes. The fraction of these manifolds is simply the volume of phase space of t and t_0 in the last term of Eq. (3). Another fraction, given by the volume of phase space in the first integral of (3) corresponds to line segments touching the margin planes at only one endpoint. The remainder of the manifolds are those interior to the margin planes. Fig. 1 shows that our theoretical calculations correspond nicely with our numerical simulations for the perceptron capacity of line segments, even with modest input dimensionality $N = 200$. Note that as $R \to \infty$, half of the

manifolds lie in the plane while half only touch it; however, the angles between these segments and the margin planes approach zero in this limit. As $R \to 0$, half of the points lie in the plane [27].

D-dimensional balls: Higher dimensional manifolds arise from multiple sources of variability and their nonlinear effects on the neural responses. An example is varying stimulus orientation, resulting in two-dimensional object manifolds under the cosine tuning function (Fig. 2(a)). Linear classification of these manifolds depends only upon the properties of their convex hulls [30]. We consider simple convex hull geometries as D-dimensional balls embedded in N-dimensions: $\left\{ \mathbf{x}^\mu + R \sum_{i=1}^D s_i \mathbf{u}_i^\mu \right\}$, so that the μ-th manifold is centered at the vector $\mathbf{x}^\mu \in \mathcal{R}^N$ and its extent is described by a set of D basis vectors $\left\{ \mathbf{u}_i^\mu \in \mathcal{R}^N, \ i = 1, ..., D \right\}$. The points in each manifold are parameterized by the D-dimensional vector $\vec{s} \in \mathcal{R}^D$ whose Euclidean norm is constrained by: $\|\vec{s}\| \leq 1$ and the radius of the balls are quantified by R.

Statistically, all components of $\mathbf{x}^\mu$ and $\mathbf{u}_i^\mu$ are i.i.d. Gaussian random variables with zero mean and unit variance. We define $h_0^\mu = N^{-1/2} y^\mu \mathbf{w} \cdot \mathbf{x}^\mu$ as the field induced by the manifold centers and $h_i^\mu = N^{-1/2} y^\mu \mathbf{w} \cdot \mathbf{u}_i^\mu$ as the D fields induced by each of the basis vectors and with normalization $\|\mathbf{w}\| = \sqrt{N}$. To classify all the points on the manifolds correctly with margin κ, $\mathbf{w} \in \mathcal{R}^N$ must satisfy the inequality $h_0^\mu - R \|\vec{h}^\mu\| \geq \kappa$ where $\|\vec{h}^\mu\|$ is the Euclidean norm of the D-dimensional vector $\vec{h}^\mu$ whose components are h_i^μ. This corresponds to the requirement that the field induced by the points on the μ-th manifold with the smallest projection on $\mathbf{w}$ be larger than the margin κ.

We solve the replica theory in the limit of $N, P \to \infty$ with finite $\alpha = P/N$, D, and R. The fields for each of the manifolds can be written as sums of Gaussian quenched and entropic components, $\left(t_0 \in \mathcal{R}, \ \vec{t} \in \mathcal{R}^D \right)$ and $\left(z_0 \in \mathcal{R}, \ \vec{z} \in \mathcal{R}^D \right)$, respectively. The capacity for D-dimensional manifolds is given by the replica symmetric calculation (SM):

$$\alpha_D^{-1}(\kappa, R) = \left\langle \min_{t_0 + z_0 - R \|\vec{t} + \vec{z}\| > \kappa} \frac{1}{2} \left[z_0^2 + \|\vec{z}\|^2 \right] \right\rangle_{t_0, \vec{t}} . \quad (4)$$

The capacity calculation can be partitioned into three regimes. For large $t_0 > \kappa + Rt$, where $t = \|\vec{t}\|$, $z_0 = 0$ and $\vec{z} = 0$ corresponding to manifolds which lie interior to the margin planes of the perceptron. On the other hand, when $t_0 < \kappa - R^{-1}t$, the minimum is obtained at $z_0 = \kappa - t_0$ and $\vec{z} = -\vec{t}$ corresponding to manifolds which are fully embedded in the margin planes. Finally, in the intermediate regime, when $\kappa - R^{-1}t < t_0 < \kappa + Rt$, $z_0 = R \|\vec{t} + \vec{z}\| - t_0 + \kappa$ but $\vec{z} \neq -\vec{t}$ indicating that these manifolds only touch the margin plane. Decomposing the capacity over these regimes and integrating out the angular components, the capacity of the perceptron can be written as:

$$\alpha_D^{-1}(\kappa, R) = \int_0^\infty dt \, \chi_D(t) \int_{\kappa - \frac{1}{R} t}^{\kappa + Rt} Dt_0 \frac{(Rt + \kappa - t_0)^2}{R^2 + 1}$$
$$+ \int_0^\infty dt \, \chi_D(t) \int_{-\infty}^{\kappa - \frac{1}{R} t} Dt_0 \left[(\kappa - t_0)^2 + t^2 \right] \quad (5)$$

where $\chi_D(t) = \frac{2^{1-\frac{D}{2}}}{\Gamma(\frac{D}{2})} t^{D-1} e^{-\frac{1}{2} t^2}$ is the D-Dimensional Chi probability density function. For large $R \to \infty$, Eq. (5) reduces to: $\alpha_D^{-1}(\kappa, \infty) = \alpha_0^{-1}(\kappa) + D$ which indicates that $\mathbf{w}$ must be in the null space of the PD basis vectors $\{\mathbf{u}_i^\mu\}$ in this limit. This case is equivalent to the classification of P points (the projections of the manifold centers) by a perceptron in the $N - PD$ dimensional null space.

To probe the fields, we consider the joint distribution of the field induced by the center, h_0, and the norm of the fields induced by the manifold directions, $h \equiv \|\vec{h}\|$. There are three contributions. The first term corresponds to $h_0 - Rh > \kappa$, i.e. balls that lie interior to the perceptron margin planes; the second component corresponds to $h_0 - Rh = \kappa$ but $h > 0$, i.e. balls that touch the margin planes; and the third contribution represents the fraction of balls obeying $h_0 = \kappa$ and $h = 0$, i.e. balls fully embedded in the margin. The dependence of these contributions on R for $D = 2$ is shown in Fig. 2(b). Interestingly, when $\kappa = 0$, the case of $R = 1$ is particularly simple for all D. The capacity is $\alpha_D = 2/(D+1)$; in addition, the fraction of embedded and interior balls are equal and the fraction of touching balls have a maximum, see Fig. 2(b) and SM.

In a number of realistic problems, the dimensionality D of the object manifolds could be quite large. Hence, we analyze the limit $D \gg 1$. In this situation, for the capacity to remain finite, R has to be small, scaling as $R \propto D^{-\frac{1}{2}}$, and the capacity is $\alpha_D(\kappa, D) \approx \alpha_0(\kappa + R\sqrt{D})$. In other words, the problem of separating P high dimensional balls with margin κ is equivalent to separating P points but with a margin $\kappa + R\sqrt{D}$. This is because when the distance of the closest point on the D-dimensional ball to the margin plane is κ, the distance of the center is $\kappa + R\sqrt{D}$ (see Fig. 2). When R is larger, the capacity vanishes as $\alpha_D(0, R) \approx \left(1 + R^{-2}\right) / D$. When D is large, making $\mathbf{w}$ orthogonal to a significant fraction of high dimensional manifolds incurs a prohibitive loss in the effective dimensionality. Hence, in this limit, the fraction of manifolds that lie in the margin plane is zero. Interestingly, when R is sufficiently large, $R \propto \sqrt{D}$, it becomes advantageous for $\mathbf{w}$ to be orthogonal to a finite fraction of the manifolds.

L_p balls: To study the effect of changing the geometrical shape of the manifolds, we replace the Euclidean norm constraint on the manifold boundary by a constraint on their L_p norm. Specfically, we consider D-dimensional manifolds $\left\{ x^\mu + R \sum_{i=1}^D s_i u_i^\mu \right\}$ where the D dimensional vector $\vec{s}$ parameterizing points on the manifolds is constrained: $\|\vec{s}\|_p \leq 1$. For $1 < p < \infty$, these L_p
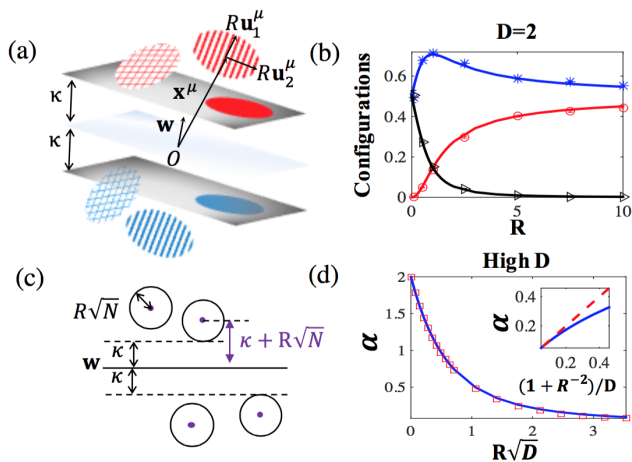
FIG. 2: Random D-dimensional balls: (a) Linear classification of $D = 2$ balls. (b) Fraction of 2-D ball configurations as a function of R at capacity with $\kappa = 0$, comparing theory (lines) with simulations (markers). (red) balls embedded in the plane, (blue) balls touching the plane, (black) interior balls. (c) Linear classification of balls with $D = N$ at margin κ (black circles) is equivalent to point classification of centers with effective margin $\kappa + R\sqrt{N}$ (purple points). (d) Capacity $\alpha = P/N$ for $\kappa = 0$ for large $D = 50$ and $R \propto D^{-1/2}$ as a function of $R\sqrt{D}$. (blue solid) $\alpha_D(0, R)$ compared with $\alpha_0(R\sqrt{D})$ (red square). (Inset) Capacity α at $\kappa = 0$ for $0.35 \leq R \leq 20$ and $D = 20$: (blue) theoretical α compared with approximate form $(1 + R^{-2})/D$ (red dashed).

manifolds are smooth and convex. Their linear classification by a vector $\mathbf{w}$ is determined by the field constraints $h_0^\mu - R||\vec{h}^\mu||_q \geq \kappa$ where, as before, h_0^μ are the fields induced by the centers, and $||\vec{h}^\mu||_q$, $q = p/(p-1)$, are the L_q dual norms of the D-dimensional fields induced by $\mathbf{u}_i^\mu$ (SM). The resultant solutions are qualitatively similar to what we observed with L_2 ball manifolds.

However, when $p \leq 1$, the convex hull of the manifold becomes faceted, consisting of vertices, flat edges and faces. For these geometries, the constraints on the fields associated with a solution vector $\mathbf{w}$ becomes: $h_0^\mu - R \max_i |h_i^\mu| \geq \kappa$ for all $p < 1$. We have solved in detail the case of $D = 2$ (SM). There are four manifold classes: interior; touching the margin plane at a single vertex point; a flat side embedded in the margin; and fully embedded. The fractions of these classes are shown in Fig. 3.

Discussion: We have extended Gardner's theory of the linear classification of isolated points to the classification of continuous manifolds. Our analysis shows how linear separability of the manifolds depends intimately upon the

dimensionality, size and shape of the convex hulls of the manifolds. Some or all of these properties are expected to differ at different stages in the sensory hierarchy. Thus, our theory enables systematic analysis of the degree to which this reformatting enhances the capacity for object classification at the higher stages of the hierarchy.
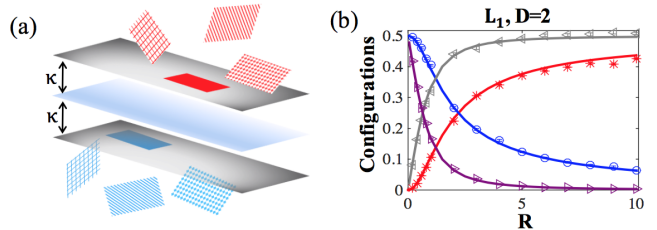

FIG. 3: L_1 balls: (a) Linear classification of 2-D L_1 balls. (b) Fraction of manifold configurations as a function of radius R at capacity with $\kappa = 0$ comparing theory (lines) to simulations (markers). (red) entire manifold embedded, (blue) manifold touching margin at a single vertex, (gray) manifold touching with two corners (one side), (purple) interior manifold.

We focused here on the classification of fully observed manifolds and have not addressed the problem of generalization from finite input sampling of the manifolds. Nevertheless, our results about the properties of maximum margin solutions can be readily utilized to estimate generalization from finite samples. The current theory can be extended in several important ways. Additional geometric features can be incorporated, such as non-uniform radii for the manifolds as well as heteogeneous mixtures of manifolds. The influence of correlations in the structure of the manifolds as well as the effect of sparse labels can also be considered. The present work lays the groundwork for a computational theory of neuronal processing of objects, providing quantitative measures for assessing the properties of representations in biological and artificial neural networks.

[1] J. J. DiCarlo and D. D. Cox, Trends in Cognitive Sciences 11, 333 (2007).

[2] M. Pagan, L. S. Urban, M. P. Wohl, and N. C. Rust, Nature Neuroscience 16, 1132 (2013).

[3] A. Alemi-Neissi, F. B. Rosselli, and D. Zoccolan, The Journal of Neuroscience 33, 5939 (2013).

[4] J. K. Bizley and Y. E. Cohen, Nature Reviews Neuroscience 14, 693 (2013).

[5] E. M. Meyers, M. Borzello, W. A. Freiwald, and D. Tsao, The Journal of Neuroscience 35, 7069 (2015).

[6] R. F. Schwarzlose, J. D. Swisher, S. Dang, and N. Kanwisher, Proceedings of the National Academy of Sciences 105, 4447 (2008).

[7] J. A. Gottfried, Nature Reviews Neuroscience 11, 628 (2010).

[8] C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo, Science 310, 863 (2005).

[9] W. A. Freiwald and D. Y. Tsao, Science 330, 845 (2010).

[10] C. F. Cadieu, H. Hong, D. L. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. Di- Carlo, PLoS Comput Biol 10, e1003963 (2014).

[11] E. Kobatake and K. Tanaka, Journal of neurophysiology 71, 856 (1994).

[12] N. C. Rust and J. J. DiCarlo, The Journal of Neuroscience 30, 12978 (2010).

[13] M. L. Minsky and S. A. Papert, Perceptrons - Expanded Edition: An Introduction to Computational Geometry (MIT press Boston, MA:, 1987).

[14] E. Gardner, Europhysics Letters 4, 481 (1987).

[15] T. Serre, L. Wolf, and T. Poggio, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR (IEEE, 2005), vol. 2, pp. 994–1000.

[16] I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, and A. Y. Ng, in Advances in Neural Information Processing Systems (2009), pp. 646–654.

[17] M. A. Ranzato, F. J. Huang, Y.-L. Boureau, and Y. Le-Cun, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR (IEEE, 2007), pp. 1–8.

[18] Y. Bengio, Foundations and Trends in Machine Learning 2, 1 (2009).

[19] E. Gardner, Journal of physics A: Mathematical and General 21, 257 (1988).

[20] A. Engel, C. Van den Broeck, and C. Broeck, Statistical Mechanics of Learning (Cambridge University Press, 2001).

[21] M. Advani, S. Lahiri, and S. Ganguli, Journal of Statistical Mechanics: Theory and Experiment 2013, P03014 (2013).

[22] N. Brunel, V. Hakim, P. Isope, J.-P. Nadal, and B. Barbour, Neuron 43, 745 (2004).

[23] R. Rubin, R. Monasson, and H. Sompolinsky, Physical Review Letters 105, 218102 (2010).

[24] H. Sompolinsky, N. Tishby, and H. S. Seung, Physical Review Letters 65, 1683 (1990).

[25] D. J. Amit, K. Wong, and C. Campbell, Journal of Physics A: Mathematical and General 22, 2039 (1989).

[26] R. Monasson, Journal of Physics A: Mathematical and General 25, 3701 (1992).

[27] L. F. Abbott and T. B. Kepler, Journal of Physics A: Mathematical and General 22, 2031 (1989).

[28] B. Lopez, M. Schroder, and M. Opper, Journal of Physics A: Mathematical and General, 28.16, L447 (1995).

[29] V. Vapnik, Statistical Learning Theory, vol. 1 (Wiley New York, 1998).

[30] M. De Berg, M. Van Kreveld, M. Overmars, and O. C. Schwarzkopf, Computational geometry (Springer, 2000).