Title

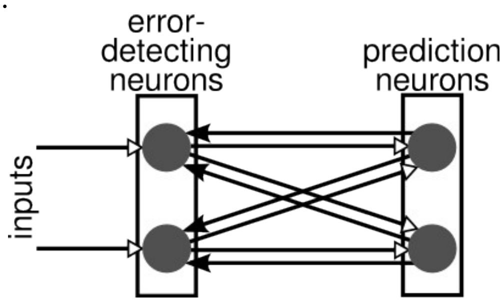
Predictive Coding

Definition

Predictive Coding is both a technique for efficient information encoding and a method for performing perceptual inference. It is commonly used to model information processing in the cerebral cortex.

Detailed Description

Predictive coding models of cortical function are typically implemented as hierarchical neural networks. Such networks contain alternating populations of "error-detecting" neurons and "prediction" neurons. The diagram shows the connectivity between a single pair of adjacent populations in such a hierarchy, for a very simplified case where there are only two neurons in each population. The inputs come from the thalamus or are the outputs of prediction neurons at preceding stages in the hierarchy. The activity of the prediction neurons encodes hypotheses about the causes underlying the inputs to the preceding population of error-detecting neurons. The activity of the error-detecting neurons encodes the discrepancy (or residual error) between the expected inputs (reconstructed from the predictions) and the actual inputs. The activations of both the prediction neurons and the error-detecting neurons are updated via an iterative process to find the combination of prediction neuron responses that best explain the inputs (in terms of minimizing the residual error), and to encode this error.



The weights from the error-detecting neurons to an individual prediction neuron determine the pattern of input activations which that particular prediction neuron represents. The strength of activation of that prediction neuron encodes the degree of belief that this specific pattern of inputs is present in the current stimulus. The full range of possible patterns that the network can represent are defined by the weights of all the prediction neurons. If these weights are represented by a matrix, $\boldsymbol{W}$, then the rows of $\boldsymbol{W}$ (which correspond to the weights targeting individual prediction neurons) comprise "basis vectors" or "elementary components" or "codebook entries" or "atoms". W as a whole represents a "dictionary" or "codebook" of possible representations, or a model of the external world. The feedback weights, connecting the prediction neurons to the error-detecting neurons, are typically chosen to be the transpose of the feedforward weights (i.e. $\boldsymbol{W}^T$). This means that the feedforward weight connecting a specific error-detecting neuron to a specific prediction neuron is equal to the reciprocal feedback weight connecting the same prediction neuron to the same error-detecting neuron. The feedback weights are used to reconstruct the expected inputs, based on the current prediction neuron responses. If prediction neuron responses are represented by the vector $\boldsymbol{y}$, then this reconstruction is:

$$\boldsymbol{r} = \boldsymbol{W}^T\boldsymbol{y} \tag{1}$$

The reconstruction thus takes the form of a linear generative model. In other words, the input is reconstructed using a linear superposition of dictionary elements where the contribution of each dictionary element is weighted by the coefficients, $\boldsymbol{y}$.

Many other algorithms employ linear generate models, and hence, are closely related to predictive coding. For example, many methods in signal processing and pattern recognition seek to approximate an input signal, $\boldsymbol{x}$, as a linear combination of basis vectors selected from a dictionary, $\boldsymbol{W}^{T}$, in proportion to a set of coefficients, $\boldsymbol{y}$. Well-known examples include vector quantization (VQ), principal components analysis (PCA), independent components analysis (ICA), and non-negative matrix factorization (NMF). In signal processing Linear Predictive Coding is used to encode a temporal sequence of signals using a linear combination of signals from the recent past, hence, the dictionary is composed of previous signals and may change through time. Sparse coding (Olshausen and Field, 1996) is also closely related, as it employs a linear generative model, but one in which y is constrained to contain only a few non-zero elements, so as to represent the input using a sparse set of coefficients. Unlike predictive coding these related methods are not hierarchical, however, algorithms that employ a hierarchy of generative models, and which are also therefore closely related to predictive coding, include the Helmholtz machine, the restricted Boltzmann machine, stacked auto-encoders, and deep learning architectures in general (Arel et al 2010).

The linear reconstruction of the inputs is compared to the actual inputs ($\boldsymbol{x}$) in order to calculate the residual error. This calculation is performed by the error-detecting neurons and is encoded in the responses of these neurons (represented by the vector $\boldsymbol{e}$). Two methods of calculating the error have been suggested. The more common method uses subtraction (Rao and Ballard, 1999), such that:

$$\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{r} \tag{2}$$

If the predictions are accurate, then the reconstructed input will equal the actual input and the error will be zero. Elements of the input vector which are not correctly predicted will result in positive or negative values in the corresponding elements of $\boldsymbol{e}$. An alternative method uses element-wise division (Spratling, 2008), such that:

$$\boldsymbol{e} = \boldsymbol{x}/\boldsymbol{r} \tag{3}$$

In this case, elements of the input that are accurately represented by the predictions will have a value of one, and elements of the input that are not correctly predicted will result in values of $\boldsymbol{e}$ greater or less than one.

The aim of predictive coding is to find the combination of prediction neuron activations that minimize the residual error. Hence, the error-detecting neuron activations are used to calculate new prediction neuron responses with the aim of improving the predictions and subsequently reducing the error. At each iteration, the prediction neuron activations are updated as follows:

$$\boldsymbol{y} \leftarrow \boldsymbol{a}\boldsymbol{y} + \boldsymbol{b}\boldsymbol{W}\boldsymbol{e} \tag{4}$$

or

$$\boldsymbol{y} \leftarrow \boldsymbol{y} \,.\, \boldsymbol{W}\boldsymbol{e} \tag{5}$$

Where equation 4 is used if equation 2 has been used to calculate the residual errors, and equation 5 is used if equation 3 has been used to calculate the residual errors. $\boldsymbol{a}$ and $\boldsymbol{b}$ are scalar parameters. In either case, errors that are large (positive values from equation 2 or values greater than one from equation 3) correspond to elements of the input that are under-represented in the reconstruction, and the above equation leads to an increase the responses of prediction neurons that represent those elements of the input. Similarly, elements of the input that are over-represented in the reconstruction, giving rise to errors that are negative (equation 2) or less-than one (equation 3) reduce the responses of prediction neurons that represent those elements of the input.

While we can think of the weights, $\boldsymbol{W}$, as defining the pattern of input activations that each prediction neuron represents, or the receptive fields (RFs) of each neuron, the prediction neurons do not behave as linear feature detectors, or linear filters, for these input patterns. If multiple prediction neurons all receive non-zero connections from the current active sub-set of inputs, then in the linear case each of those prediction neurons would respond in proportion to the overlap between the stimulus and that neuron's RF (i.e. $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$). In contrast, in predictive coding the activation dynamics described above cause the prediction neurons to interact in a mutually suppressive and nonlinear manner to determine the sub-set of those predictions which can best explain the input pattern. This interactive process can be thought of as a form of perceptual inference in which the prediction neuron responses represent hypotheses and the inputs provide evidence in support of those hypotheses. If a prediction neuron (or sub-set of prediction neurons) are able to explain the input, then the evidence supporting alternative hypotheses is “explained away” (Lochmann and Deneve, 2011) and the the prediction neurons representing those alternative hypotheses do not respond. This perceptual inference process can also be thought of in terms for competition between the prediction neurons. Each prediction neuron effectively tries to block other prediction neurons from responding to the inputs which it represents, such that prediction neurons with overlapping RFs compete for the right to represent active inputs.

If the input remains constant, at the end of several iterations of equations 1, 2 and 4 or of equations 1, 3, and 5 the prediction neuron responses will typically reach a steady-state where the active prediction neurons represent the linear combination of dictionary elements that can best explain the input. In the longer-term the dictionary elements (i.e. the weights $\boldsymbol{W}$) can be adjusted to provide a better set of elementary components with which to encode the input space. Like the updating of the prediction neuron responses, this learning process is also error-driven, such that:

$\Delta \boldsymbol{W} = fn(\boldsymbol{ye})$ if equation (2) has been used to calculate the residual errors

or

$\Delta \boldsymbol{W} = fn(\boldsymbol{y}(\boldsymbol{e}-1))$ if equation (3) has been used to calculate the residual errors.

The above description has been concerned with the interactions between a single pair of error-detecting and prediction neuron populations. In a hierarchical predictive coding model there will be multiple such pairs, and the activity in one pair will be influenced by the subsequent pair, higher up the hierarchy. For example,
equation 4 will become $\boldsymbol{y} \leftarrow a\boldsymbol{y}+b\boldsymbol{We}-c\boldsymbol{e}^*$ where $\boldsymbol{e}^*$ is the activation of the subsequent population of error-detecting neurons in the hierarchy, and c is a scalar parameter.
Alternatively,
equation 5 will become $\boldsymbol{y} \leftarrow \boldsymbol{y}.\boldsymbol{We} + d\boldsymbol{W}^{*T}\boldsymbol{y}^*$ where $\boldsymbol{W}^{*T}$ and $\boldsymbol{y}^*$ are the weights and prediction neuron responses of the subsequent stage in the hierarchy, and d is a scalar parameter. In either case prediction neuron responses at one stage in the hierarchy can enhance the responses of neurons representing corresponding predictions at the preceding stage.

Predictive coding is an abstract, functional, model which explores the computational principles by which the cortex operates, rather than the biophysical mechanisms by which these computations may be implemented (Spratling, 2011). It is therefore difficult to relate specific aspects of the model to individual components of cortical physiology. However, the most common interpretation is that cortical feedback connections carry predictions and that these act on regions at preceding stages along an information processing pathway in order to calculate the residual error which is then propagated via cortical feedforward connections (Rao and Ballard, 1999; Friston, 2005). Under this interpretation, pyramidal cells in the superficial layers of primary visual cortex (V1) would correspond to error-detecting neurons and some aspects of V1 complex cell behavior can be modeled by error-detecting neurons in a suitable predictive coding model (Rao and Ballard, 1999). An alternative interpretation proposes that the calculation of the residual error is performed by connections intrinsic to each cortical region, and hence that both cortical feedforward and feedback pathways carry predictions (Spratling, 2008). Under this interpretation, pyramidal cells in the superficial layers of primary visual cortex would correspond to prediction neurons. Many properties of orientation-tuned cells in V1 can be simulated by the prediction neurons in a suitable predictive coding model (Spratling, 2010; 2011).

Predictive coding has also been proposed as a model of retinal information processing (Srinivasan et al., 1982). Retinal circuitry is believed to predict the local intensity values expected at a particular spatial location (as the average intensity of all neighboring locations) and to subtract this prediction from the actual intensity. The signal generated following subtraction of the predicted information has a smaller dynamic range than the raw intensity values, and hence, can be transmitted with greater accuracy using a limited range of firing rates. By removing predictable information for the transmitted data the retina can be considered to perform redundancy reduction (the removal of statistical dependencies between neighboring pixel values) as proposed by Barlow (2001).

References

Arel, I., Rose, D. C., and Karnowski, T. P. (2010). Deep machine learning: A new frontier in artificial intelligence research. IEEE Computational Intelligence Magazine, 5(4):13–8.

Barlow, H. (2001). Redundancy reduction revisited. Network: Computation in Neural Systems, 12:241–53.

Friston, K. J. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456):815–36.

Lochmann, T. and Deneve, S. (2011). Neural processing as causal inference. Current Opinion in Neurobiology, 21(5):774– 81.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive properties by learning sparse code for natural images. Nature, 381:607–9.

Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1):79–87.

Spratling, M. W. (2008). Predictive Coding as a Model of Biased Competition in Visual Attention. Vision Research, 48(12):1391–408.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. Journal of Neuroscience, 30(9):3531–43.

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. Vision Research, 51(6):563-76.

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. Proceedings of the Royal Society of London. Series B, Biological Sciences, 216(1205):427–59.

Further Reading

Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, doi:10.1017/S0140525X12000477.

Egner, T., Monti, J. M. and Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream . The Journal of Neuroscience, 30(49):16601-8 doi: 10.1523/JNEUROSCI.2770-10.2010

Kveraga, K., Ghuman, A. S., and Bar, M. (2007). Top-down predictions in the cognitive brain. Brain and Cognition, 65:145–68.

Huang, Y. and Rao, R. P. N. (2011), Predictive coding. WIREs Cognitive Science, 2: 580–93. doi: 10.1002/wcs.142

Spratling, M. W. (2012) Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. Neural Computation, 24(1): 60-103.