

Part B Project Progress

Aim

The initial aim of the project is to investigate designing biologically plausible neural networks to solve the blind signal separation problem.

Blind signal separation problem

Suppose we have some data matrix, $S \in \mathbb{R}^{d \times N}$ where d is the dimension of each data point and N is the number of data points. We also have a mixing matrix $A \in \mathbb{R}^{m \times d}$ where m is the number of mixtures. Now we have $X = AS$. The goal of blind signal separation is to find the original data S given the mixtures X .

Although there are many models that solve this problem, they often make biologically unrealistic assumptions. We aim to build on existing works [1], keeping accuracy increasing plausibility.

Approaches

The first idea I explored was a two-layer linear network with an input layer of m mixtures and an output layer of d neurons. This model aims to decorrelate the neuron activations in the second layer, hoping to achieve signal separation.

When training, first a forward pass results in $\nu = Wx$. Then a relaxation step is performed, updating the activities of the neurons in the second layer z , minimising the energy defined by

$$E_i = \frac{1}{2} \left[(z_i - \nu_i)^2 + \sum_{j \neq i} \langle z_i z_j \rangle^2 \right].$$

After activity relaxation, we update the weights to minimise this energy. Finally, lateral connections are updated to keep an estimate of the correlations – these are also used for the activity relaxation.

$$L_{ij} \leftarrow \alpha z_i z_j + (1 - \alpha) L_{ij} \text{ for } i \neq j$$

Running experiments with this model, the network achieved decorrelation of neurons in the output layer.

I ran experiments, drawing data from d independent uniform distributions and using random mixing matrices where each entry was distributed with a standard normal. A training set of N such datapoints was curated and the training looped over this data many times.

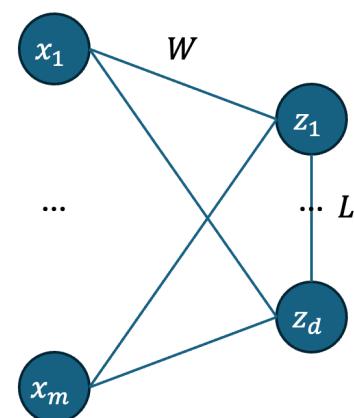


Figure 1 Two-layer neural network with lateral connections

Some data is shown for $d = 2$, $m = 2$ and $d = 3$, $m = 3$ below.

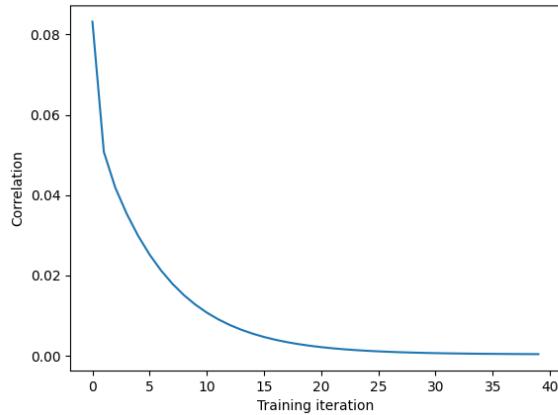


Figure 2 Correlation of activations across training for a single

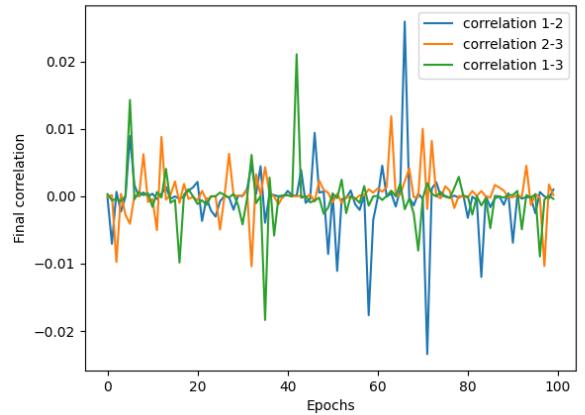


Figure 3 Mean final correlation (after training) of activations across 100 mixing matrices

However, although the activations decorrelate, the model is not very good at blind signal separation. As shown below, the model is unable to find the correct rotation to match the sources.

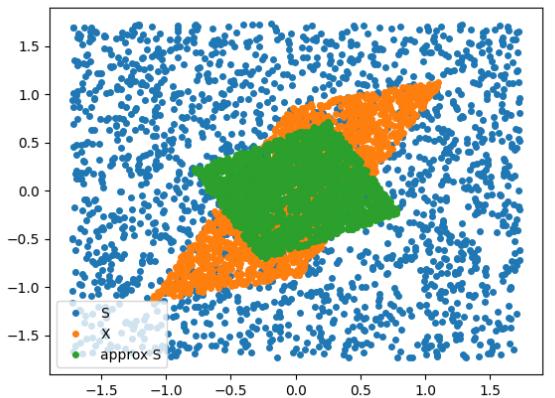
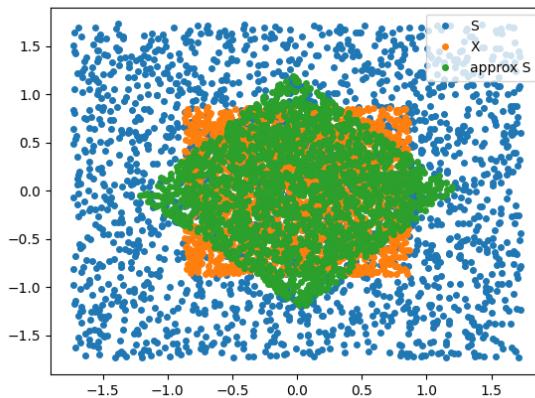


Figure 4 Examples in 2D of the model's estimated sources and the actual sources, S and mixtures, X

To resolve the correct rotation, I attempted to modify the energy function. After additionally introducing more hyperparameters, the energy function is

$$E_i = \frac{1}{2} [A(z_i - v_i)^2 + \frac{B}{d-1} \sum_{j \neq i} \langle x_i^2 x_j^2 \rangle^2 + C(1 - \text{var}(x_i))^2]$$

We also now store estimates of the variance (alongside the covariance)

$$V_i \leftarrow \alpha z_i^2 + (1 - \alpha) V_i \text{ for } i \in \{1, \dots, d\}$$

$$L_{ij} \leftarrow \alpha z_i z_j + (1 - \alpha) L_{ij} \text{ for } i \neq j$$

Now, taking gradients of the energy for activity relaxation, we get

$$\frac{\partial E_i}{\partial z_i} = A(z_i - v_i) + \frac{B}{d-1} \sum_{j \neq i} L_{ij} (\alpha z_j) + C(1 - V_i)(-2\alpha z_i)$$

Drawing inspiration from the VICREG paper [2] I also implemented a network with a similar loss function in PyTorch with batch training, estimating the covariance and variance of the activities from the batch. This is aimed to be a model that can be used to quickly test ideas and modifications to the loss function, but it is not biologically plausible.

I am currently exploring how far these variance modifications and other ideas such as adding activation functions to the second layer can go in resolving the correct rotation and solve the blind signal separation problem.

Works Cited

- [1] A. I. C. P. A. T. E. Bariscan Bozkurt, “Correlative Information Maximization Based Biologically Plausible Neural Networks for Correlated Source Separation,” in *ICLR* 2023.
- [2] J. P. Y. L. Adrien Bardes, “VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning,” in *ICLR* 2022.