ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised and reinforcement learning in neural networks
Lecture 2

Lecturer: Marc-Oliver Gewaltig

PCA & Oja's rule

Exercise 1

By looking at the formulation of the following learning rules, classify them to Hebbian-type or non-Hebbian rules.

Variable definitions. x: stimulus (neuronal input), y: neuronal activity, w: synaptic weight, ϵ: small positive number (learning rate), r : reward. Unless otherwise stated, x,y are mean firing rates.

1. Rescorla-Wanger (reward predicting) rule. $w \to w + \epsilon\delta x$, where $\delta = r - y$, $x \in \{0, 1\}$.
2. Homeostatic rule. $w \to w + \epsilon\,(y - \theta)$, where θ is the homeostatic threshold.
3. Node perturbation. $w \to w + \epsilon r y \xi$, where ξ represents noise (random variable drawn from a Gaussian distribution).
4. Sejnowski rule. $w \to w + \epsilon\,(y - \langle y\rangle)\,(x - \langle x\rangle)$, with $\langle . \rangle$ being the mean activity of the signals y and x across time.
5. Oja rule. $w \to w + \epsilon y\,(x - yw)$.
6. Learning in Hopfield networks. $w \to w + \epsilon yx$, with $x, y \in \{-1, 1\}$.
7. Delta rule. $w \to w + \epsilon\delta x$, where $\delta = t - y$, with t being the target for input x.
8. Associative Reward Inaction $w \to w + \epsilon r\,(y - P(y))\,x$, where $P(y)$ the probability of y to be active, $y \in \{0, 1\}$.

Exercise 2

Use principal component analysis to reduce the dimensionality of the dataset shown in Fig. 1.

X	3	3	4	4	5	5	6	6	7	7	8	8
Y	3	4	3	5	4	5	6	7	6	8	7	8

1. Center the data by subtracting their mean.
2. Calculate the covariance matrix of the data.

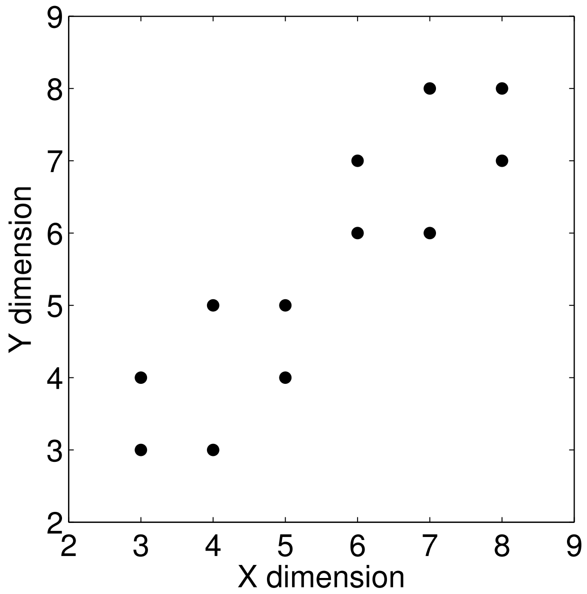

Figure 1: Original Dataset

3. Find the eigenvalues and eigenvectors of the covariance matrix and explain their meaning in the context of PCA.

4. Calculate the output data of PCA and discard the less significant component. What are the principal axes in the original coordinate system? Could you obtain the new dataset without making any calculations?

5. Can you recover the original data? How?

Exercise 3: Linear differential equations and Hebbian learning

For linear neurons, the Hebbian rule $\frac{dw}{dt} = yx$ can be written in the form $\frac{dw}{dt} = Cw$, where $C = xx^T$ is a matrix and $y = w \cdot x$ denotes the output of the neuron.

3.1 Let us consider a neuron that receives an N-dimensional input. Its weight dynamics are given by:

$$\frac{d\vec{w}}{dt} = C\vec{w} \tag{1}$$

with

$$C = \begin{pmatrix} 1 & 0.5 & 0 & 0 & \dots & 0 & 0.5 \\ 0.5 & 1 & 0.5 & 0 & \dots & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.5 & 0 & 0 & 0 & \dots & 0.5 & 1 \end{pmatrix}. \tag{2}$$

Show that the complex vectors $w_k = \exp\left(\frac{2\pi ik}{N}m\right)$, with $k = 1\dots N$ and $m \in \mathbb{Z}$ are eigenvectors of C. Assume cyclic boundary conditions.

3.2 Assume that the neuron receives N input patterns $\vec{\xi}^{\mu} = (\xi_1^{\mu}, \xi_2^{\mu}, \ldots, \xi_N^{\mu})^T$ with $\xi_k^{\mu} = \sqrt{\frac{N}{2}}\left(\delta_k^{\mu} + \delta_k^{(\mu \ mod \ N)+1}\right)$. Here, δ_k^{μ} denotes the Kronecker symbol, which is 1 if $\mu = k$ and 0 otherwise. Show that the matrix C is produced by:

$$C_{kj} = \left\langle \xi_k^{\mu} \xi_j^{\mu} \right\rangle = \frac{1}{N} \sum_{\mu=1}^{N} \xi_k^{\mu} \xi_j^{\mu}. \tag{3}$$

Comment on how the weights will evolve given the nature of the input patterns.

Exercise 4: Oja's rule extracts the first principal component

According to Oja's rule, the evolution of the synaptic weights $w(t)$ is given by:

$$\frac{d}{dt} w = Cw - (w^T C w) w. \tag{4}$$

4.1 Show that the fixed points of this equation are eigenvectors of the C matrix.

4.2 Show that the eigenvector e_k associated with the largest eigenvalue of C is a stable fixed point.

Hint: Assume that the weight is almost the eigenvector e_k, but slightly perturbed in the direction of a different eigenvector e_j: $w(t) = \alpha(t) e_k + \epsilon(t) e_j$, with $\epsilon \ll 1$ and $\epsilon^2 + \alpha^2 = 1$. Plug this ansatz into Oja's rule and show that the dynamics of $\epsilon(t)$ are given by:

$$\frac{d}{dt} \epsilon = -(\lambda_k - \lambda_j)(\epsilon - \epsilon^3)\,. \tag{5}$$

Use this result to discuss the stability of the fixed point e_k.