ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised and reinforcement learning in neural networks
Week 5

Lecturer: Marc-Oliver Gewaltig

Assistants: Dane Corneil
Berat Denizdurduran

Corrections: Competitive Learning

Exercise 1

1.1 Convergence means $\Delta\vec{w}_i = \vec{0}$. So if we have convergence, we must have

$$\eta \sum_{\mu \in C_i} (\vec{x}_\mu - \vec{w}_i) = \vec{0} \rightarrow \sum_{\mu \in C_i} \vec{x}_\mu - N\vec{w}_i = \vec{0} \rightarrow \vec{w}_i = \frac{1}{N} \sum_{\mu \in C_i} \vec{x}_\mu.$$

This is the definition of the "center of mass" of group i.

1.2 For the steady state, we know that $\langle \Delta\vec{w}_i \rangle = 0$. The variance is therefore given by:

$$\begin{aligned} Var(\Delta\vec{w}_i) &= \langle (\Delta\vec{w}_i)^2 \rangle_\mu = \langle \eta^2 (\vec{x}^\mu - \vec{w}_i))^2 \rangle_\mu \\ &= \eta^2 \langle (\vec{x}^\mu - \bar{x})^2 \rangle_\mu = \eta^2 Var(\vec{x}). \end{aligned}$$

We see that the fluctuations are proportional to the variance of the associated data points.

Exercise 2

2.1 We start by extending the squares on both sides of the inequality:

$$\vec{x}^2 - 2\vec{x}^T\vec{w}_i + \vec{w}_i^{\,2} \leq \vec{x}^2 - 2\vec{x}^T\vec{w}_k + \vec{w}_k^{\,2}.$$

Since the weight vectors are normalized, we have $\vec{w}_k^{\,2} = 1, \forall k$, the corresponding terms on each side of the inequality. We can also eliminate the $\vec{x}^2$ that appear on each side. Dividing each side by -2 (without forgetting that dividing by a negative number causes the inequality to flip direction), we get

$$\vec{x}^T\vec{w}_i \geq \vec{x}^T\vec{w}_k.$$

This means that indeed, the nearest prototype is the one with the strongest input.

2.2 Looking at the derivation of 3.1. in isolation, the argument does not hold anymore when the weight vectors are not normalized, since the scalar product with a very large weight vector in an arbitrary direction could always be the largest, irrespective of its relation to the data samples. However, from exercise 1 we know, that if we are at the steady state (or close to it), the weight vector will be the center-of-mass of the cluster it represents and hence will be approximately normalized when the data samples are:

$$\begin{aligned} \mid \vec{w} \mid^2 &= \mid \frac{1}{N} \sum \vec{x}^\mu \mid^2 = \frac{1}{N^2} (\sum_{\mu,\nu} \vec{x}^{\mu,T} \vec{x}^\nu) = \frac{1}{N^2} (\sum_\mu \vec{x}^{\mu,T}\vec{x}^\mu + \sum_{\mu,\nu,\mu\neq\nu} \vec{x}^{\mu,T}\vec{x}^\nu) \\ &= \frac{1}{N^2} (N + \sum_{\mu,\nu,\mu\neq\nu} \vec{x}^{\mu,T}\vec{x}^\nu) \approx \frac{1}{N^2} (N + \sum_{\mu,\nu,\mu\neq\nu} 1) = \frac{1}{N^2} (N + N(N-1)) = 1, \end{aligned}$$

where we assumed that the scalar product between data points *belonging to the same cluster* are close to 1 and the data samples themselves are normalized.

Exercise 3

3.1 From the lecture, we know that the algorithm we have to perform is the following:

```
for x⃗ in {x⃗1, x⃗2, x⃗1, x⃗2} do
    i* = argmax_i x⃗^T w⃗_i                  Find which neuron is activated
    Δw_i* = η(x⃗ − w⃗_i*)                    Compute the weight change
    w⃗_i* → w⃗_i* + Δw_i*                   Update the weight vector
end
```

This gives us the following sequence:

step	point	y_1	y_2	i^*	Δw_{i^*}	w_1	w_2
0	–	–	–	–	–	(1.50, 0.00, 0.50)	(0.00, 0.50, 1.50)
1	1	0.00	0.50	2	(0.00, 0.25, −0.75)	(1.50, 0.00, 0.50)	(0.00, 0.75, 0.75)
2	2	2.00	0.75	1	(−0.25, 0.00, 0.25)	(1.25, 0.00, 0.75)	(0.00, 0.75, 0.75)
3	3	0.75	1.50	2	(0.00, 0.12, 0.12)	(1.25, 0.00, 0.75)	(0.00, 0.88, 0.88)
4	1	0.00	0.88	2	(0.00, 0.06, −0.44)	(1.25, 0.00, 0.75)	(0.00, 0.94, 0.44)
5	2	2.00	0.44	1	(−0.12, 0.00, 0.12)	(1.12, 0.00, 0.88)	(0.00, 0.94, 0.44)
6	3	0.88	1.38	2	(0.00, 0.03, 0.28)	(1.12, 0.00, 0.88)	(0.00, 0.97, 0.72)

We see that $\vec{w}_1 \to \vec{x}_2$ and the second weight is going to the center of mass of $\vec{x}_1$ and $\vec{x}_3$, i. e. $\vec{w}_2 \to \frac{\vec{x}_1+\vec{x}_3}{2}$.

3.2 With only two input vectors and two units, each unit will become the prototype of exactly one input pattern. The assignment will depend on the initial conditions. In this case, $\vec{w}_1 \to \vec{x}_2$ and $\vec{w}_2 \to \vec{x}_1$.

Explicit calculation would confirm our intuition:

step	point	y_1	y_2	i^*	Δw_{i^*}	w_1	w_2
0	–	–	–	–	–	(1.50, 0.00, 0.50)	(0.00, 0.50, 1.50)
1	1	0.00	0.50	2	(0.00, 0.25, −0.75)	(1.50, 0.00, 0.50)	(0.00, 0.75, 0.75)
2	2	2.00	0.75	1	(−0.25, 0.00, 0.25)	(1.25, 0.00, 0.75)	(0.00, 0.75, 0.75)
3	1	0.00	0.75	2	(0.00, 0.12, −0.38)	(1.25, 0.00, 0.75)	(0.00, 0.88, 0.38)
4	2	2.00	0.38	1	(−0.12, 0.00, 0.12)	(1.12, 0.00, 0.88)	(0.00, 0.88, 0.38)