

# The Multi-layer Perceptron

1. What is the difference between a batch and an online update? What are the reason to prefer one over the other?
2. Why do we need to substitute the step function with the sigmoid for the Multi-Layer Perceptron (MLP)?
3. What optimization algorithm do we use, both for the perceptron and the MLP?
4. Given the dataset:  $\langle 1, 1, 0 \rangle, \langle 2, 1, 0 \rangle, \langle 1, 3, 1 \rangle, \langle 2, -1, 1 \rangle$ , where the last element of each vector is the class, construct an MLP that separates the classes.
5. Given the dataset:  $\langle 2, 1, 0 \rangle, \langle 1, 2, 0 \rangle, \langle -1, 2, 1 \rangle, \langle -1, -3, 1 \rangle, \langle 2, 3, 1 \rangle$ , construct an MLP that separates the classes.