Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447/625-438
Dr. Mark Fleischer

Module 7.3: Training with Multiple I/O Pairs

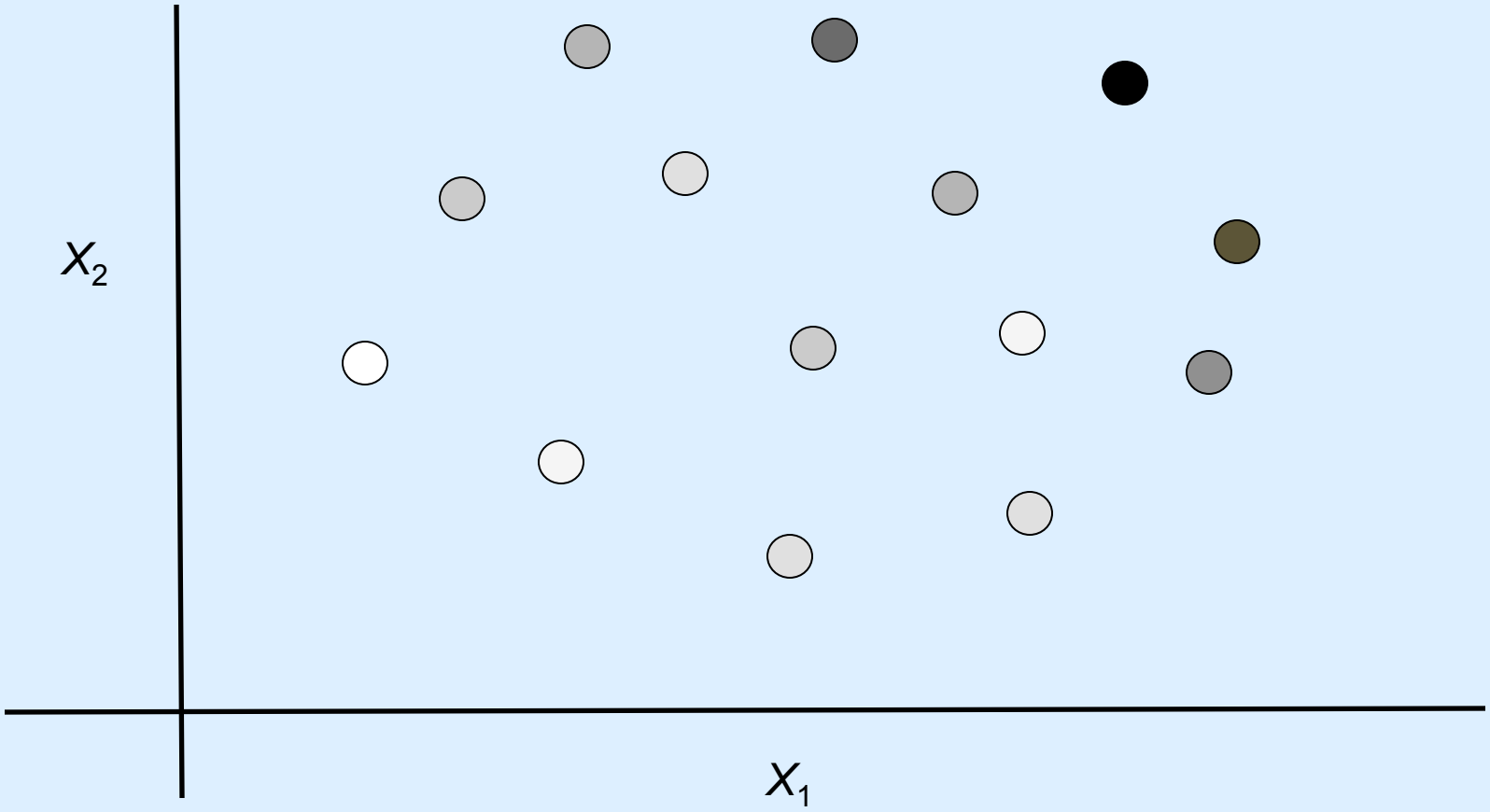
This Sub-Module Covers …

- What a typical training scenario looks like when there are multiple input/output pairs.
  - "online" or "incremental" training
  - "batch" training
- Implications and why they are different.

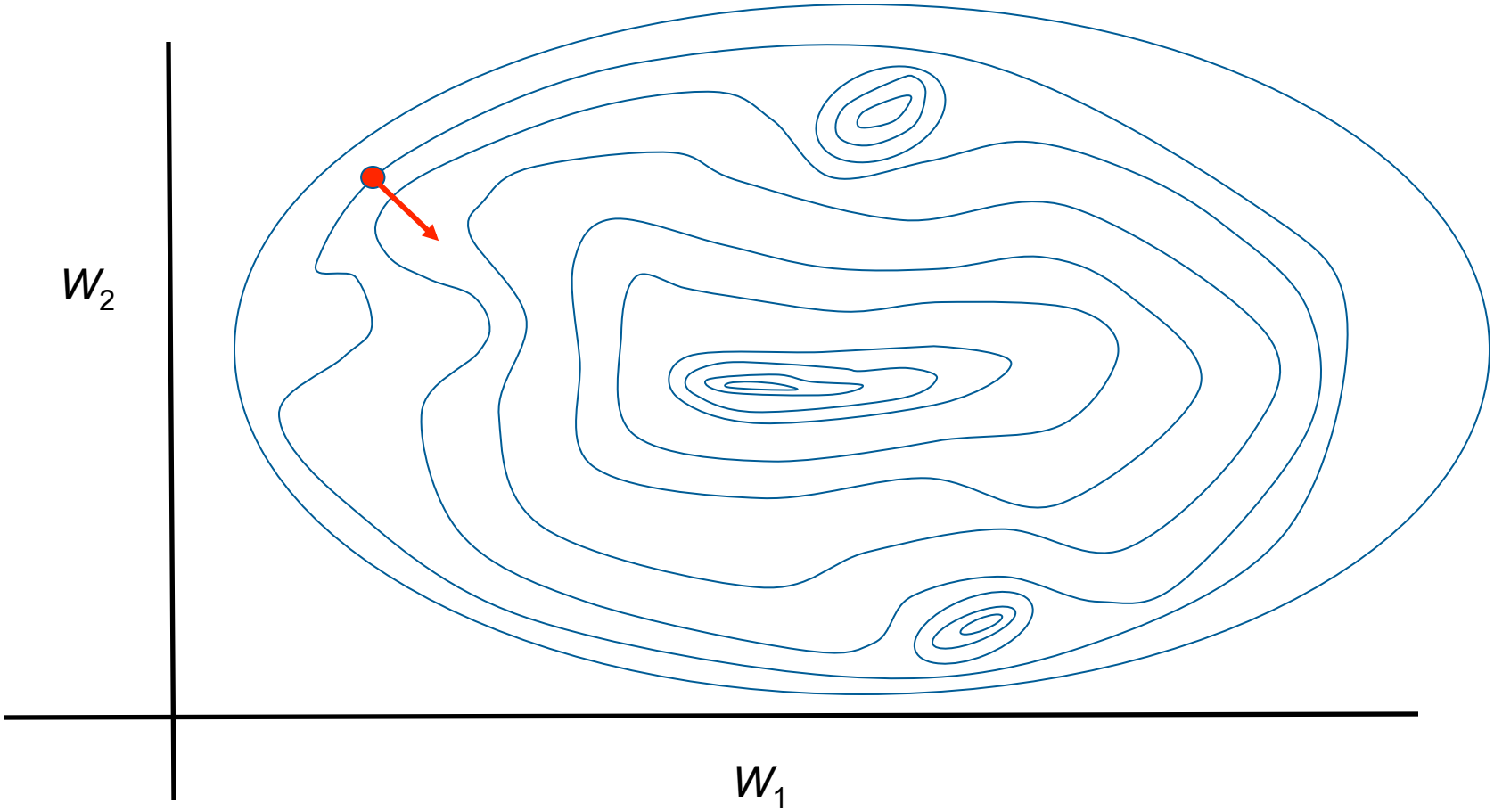
Training with Multiple Input/Output Pairs

- We've seen we can decrease errors with a single I/O pair with the Perceptron.
- Can do the same with multilayer networks.
- Still, with FFBP algorithm, training can get stuck in local optima---with one I/O pair we can get error as low as desired.
- Let's examine a type of problem with many I/O pairs.

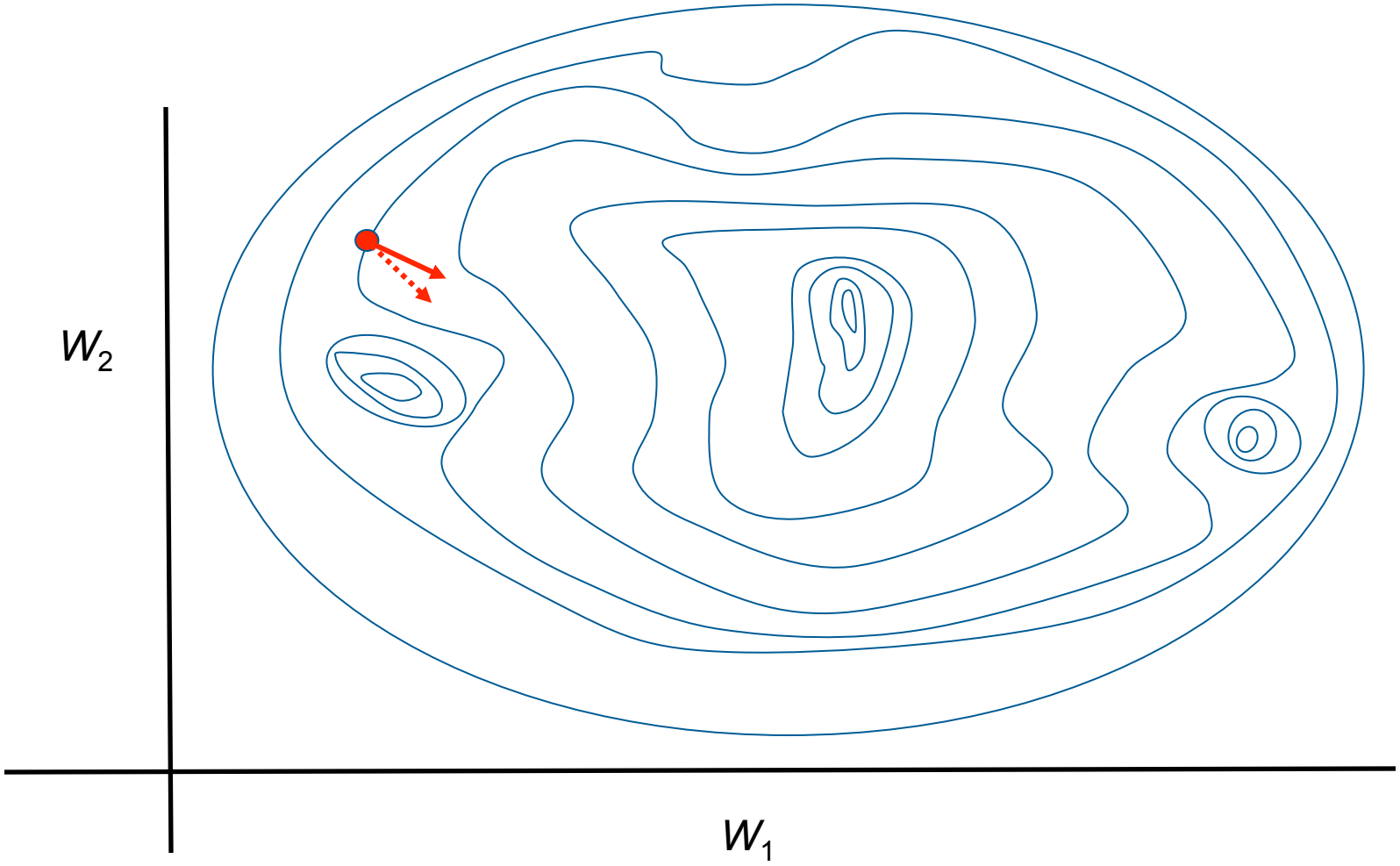
A Classification Problem



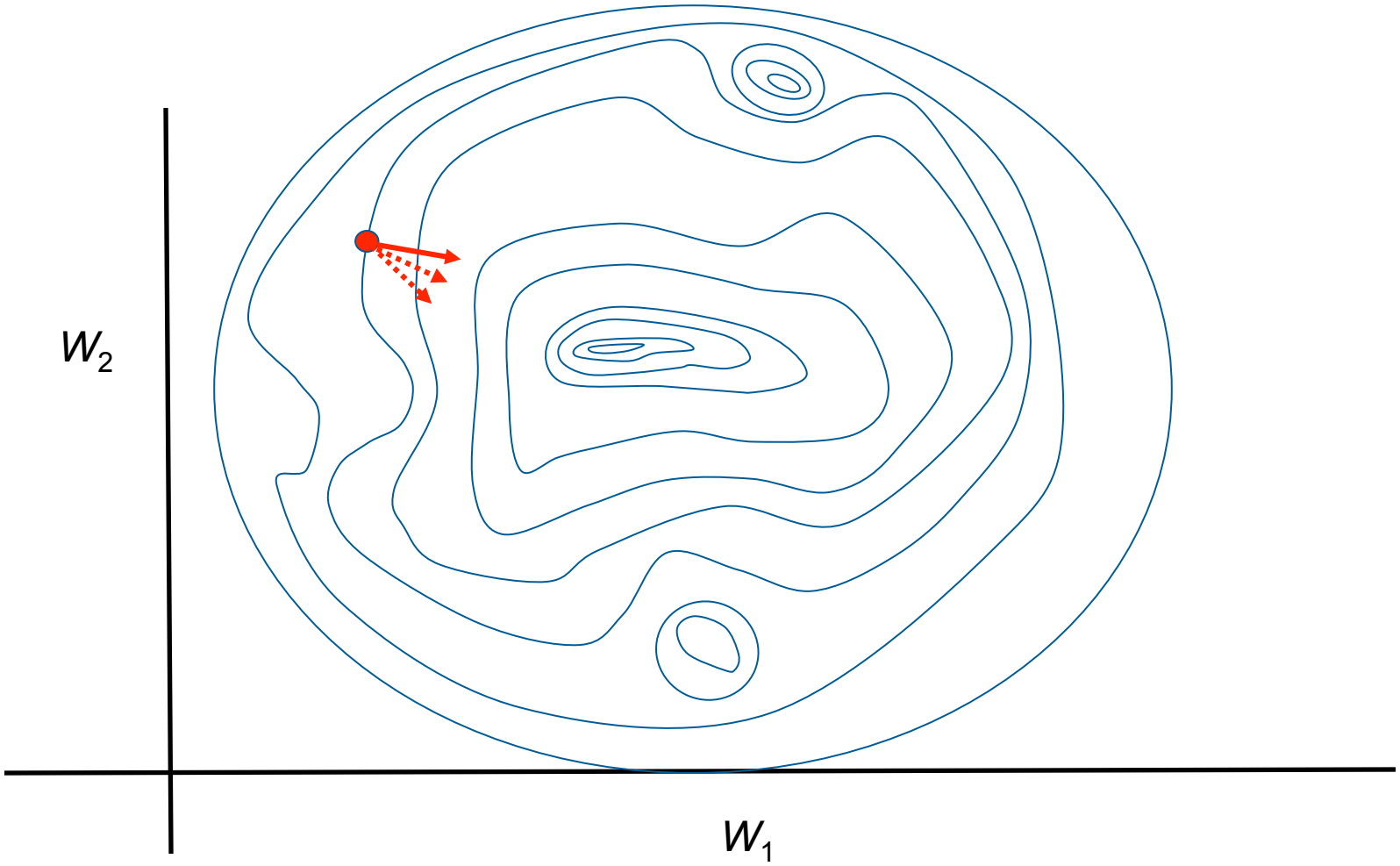
Present 1st Input/Output Pair



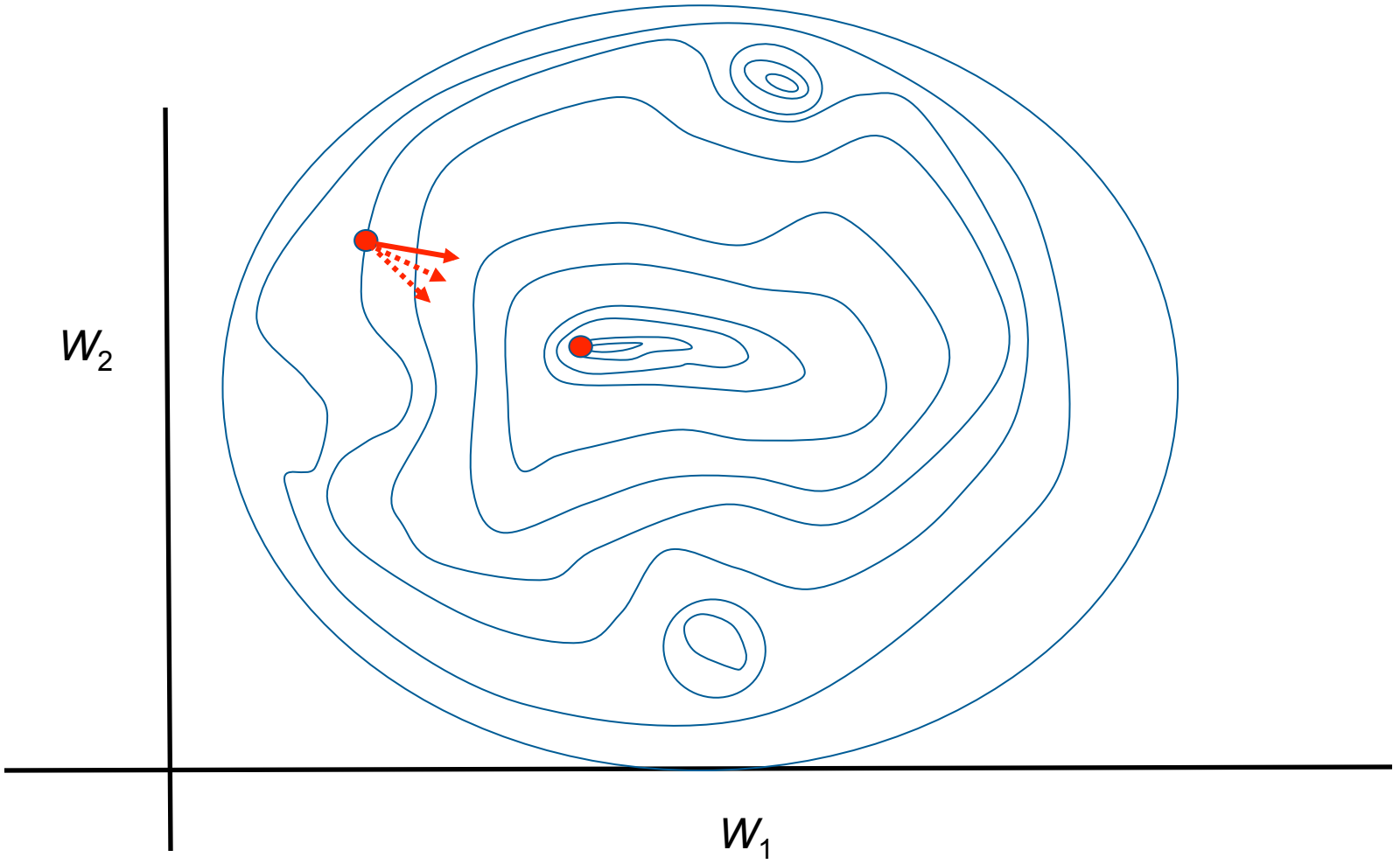
Present 2nd Input/Output Pair



Present 3rd Input/Output Pair



Updating Weights



Batch Training

Training cycle (c = 1 to C){

$\Delta w == 0$ // Initialize weight update

Present I/O pair (p = 1 to P) {

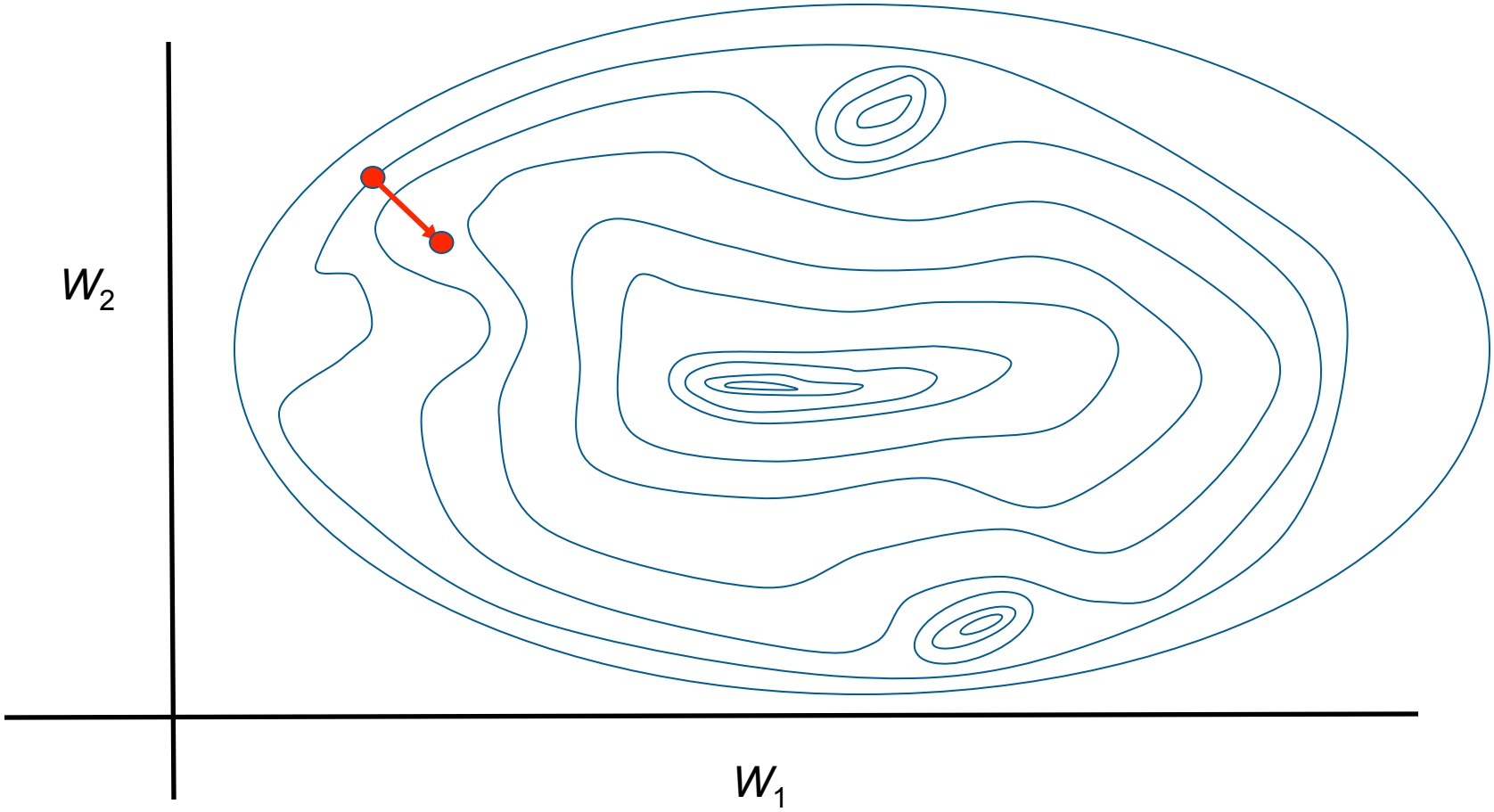
Δw += Calculate Gradient estimate();

} // Present next I/O pair

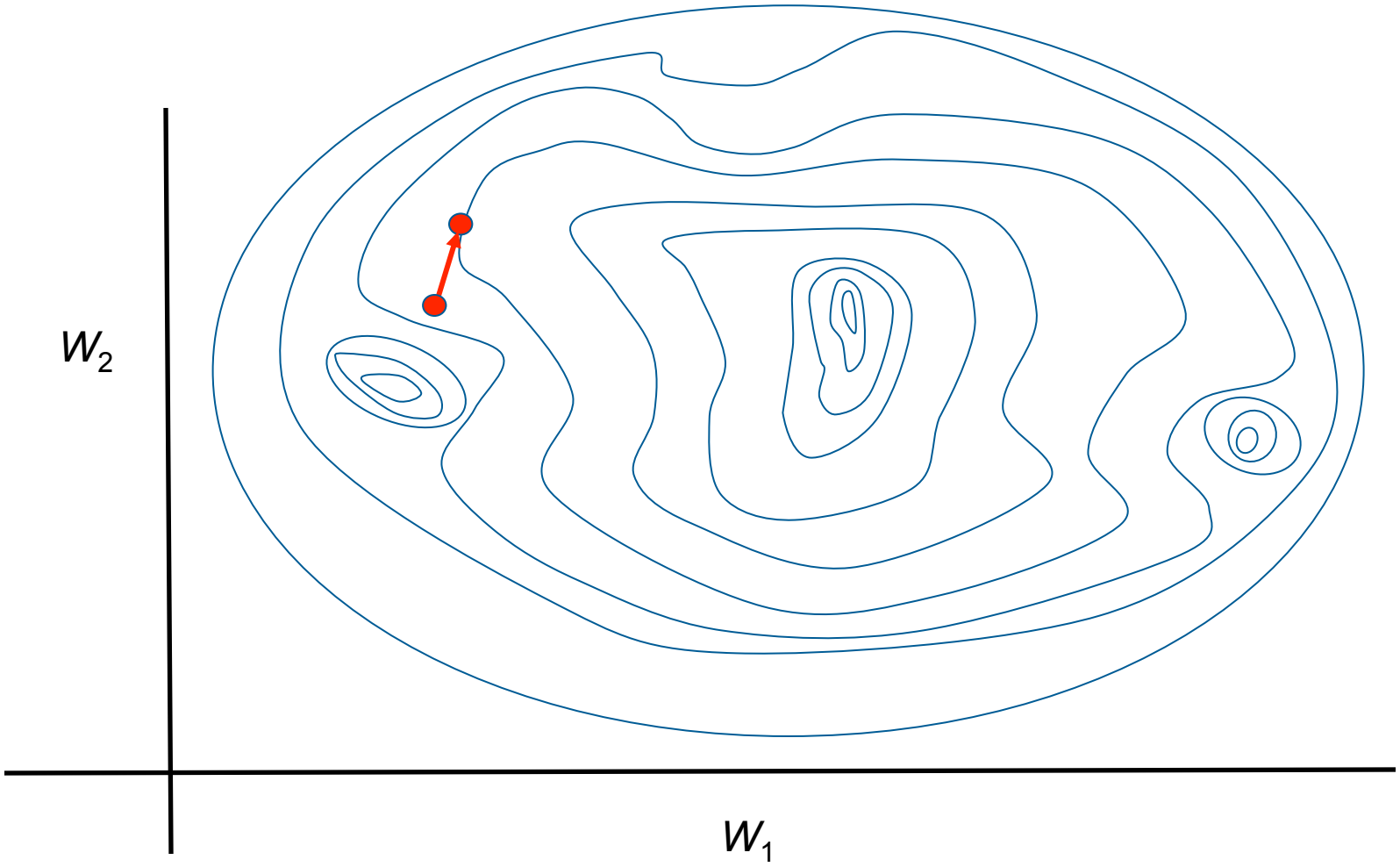
Update weight $w = w + \Delta w$

} // Next training cycle

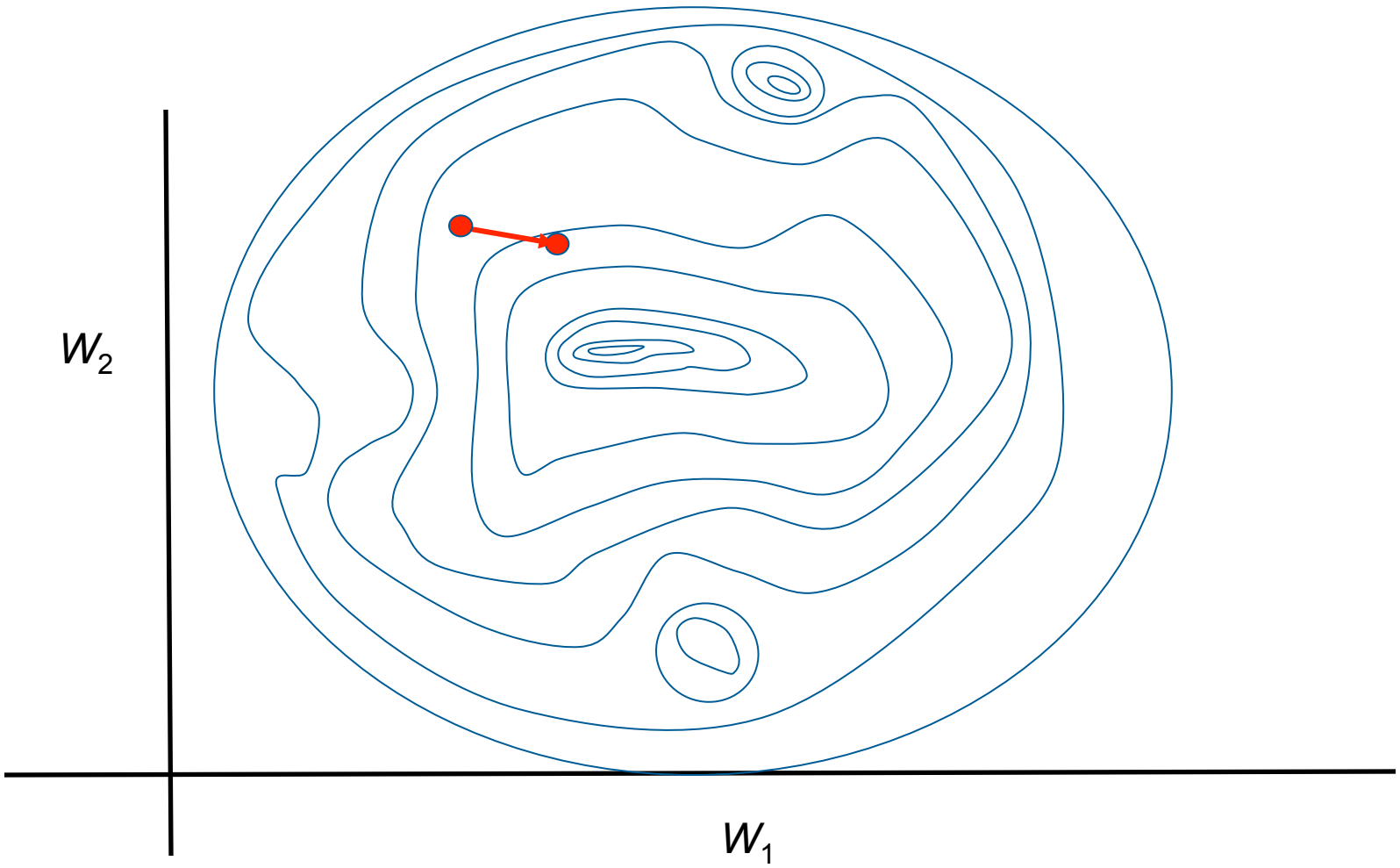
Present 1st Input/Output Pair



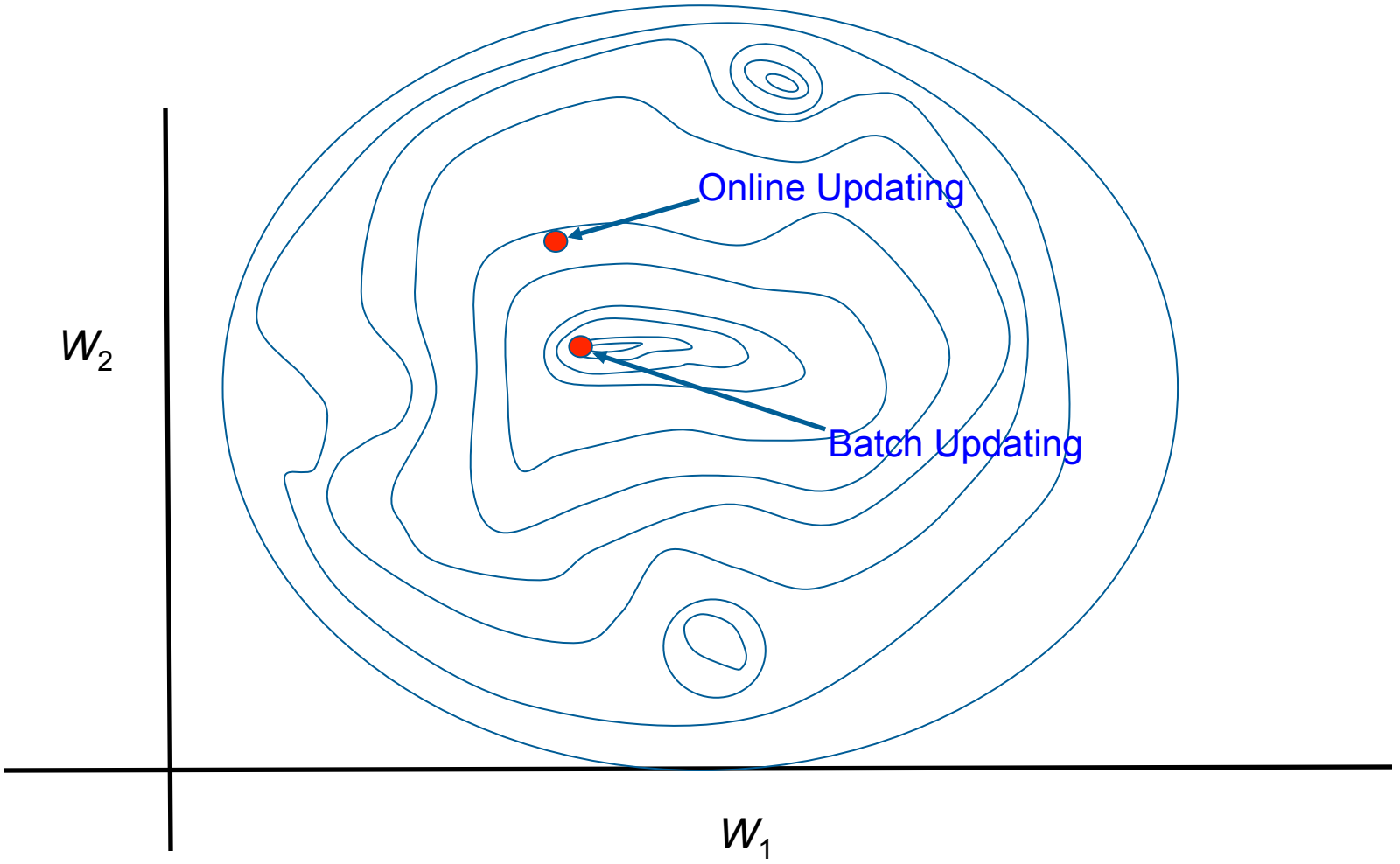
Present 2nd Input/Output Pair



Present 3rd Input/Output Pair



Comparing Updated Weights



Online (Incremental) Training

Training cycle (c = 1 to C){

Present I/O pair (p = 1 to P) {

Δw = Calculate Gradient estimate();

Update weight w = w + Δw

} // Present next I/O pair

} // Next training cycle

Batch vs. Online Training

Batch

Pros:

- Closer match to theoretical (true) gradient.
- Lower residual error.
- Permits other, useful modifications of FFBP to be used.

Online

Pros:

- Faster, earlier updates (but remember, each update changes the error function for subsequent I/O pairs).
- Useful if no ‘fixed’ training set exists or when the data are ‘nonstationary’.
- Creates effective/useful ‘noise’ that can overcome local minima.

Batch vs. Online Training

Batch

Cons:

- Requires some memory overhead.
- Can require more computation effort to get to comparable weights than online (requires all I/O pairs before an update occurs)

Online

Cons:

- Each update alters the error function landscape --- not the true gradient descent vector for all I/O pairs– perturbs the learning process.
- Sometimes noise isn't a good thing.

The Best of Both Worlds

- Use ‘online’ training for ‘initial training’ with a first part of the training data
  - Get a good, fast start with some updated weights.
- Then as we get closer to minimal error, resort to batch training
  - More carefully and accurate zero-in to true minimal error.