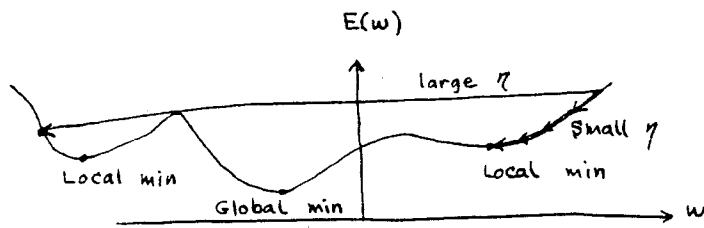


Apr 1990

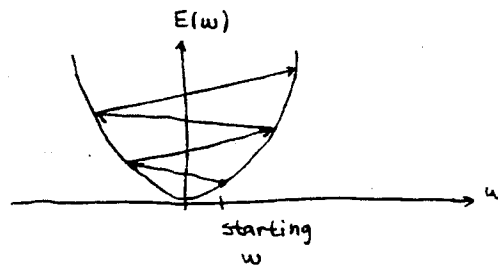
Neil E Potter

## Gradient Descent - Local Minima



If gradient descent step size,  $\eta$ , is small,  $w$  will be trapped in the bowl it starts in. Thus,  $w$  approaches a local minimum instead of the global minimum.

If  $\eta$  is too large,  $w$  will jump over the global min. Even worse, large  $\eta$  may make the algorithm unstable,  $w \rightarrow \infty$ .



There is no magical formula for choosing  $\eta$ .

Q. How do we avoid local minima?

A. Add noise to  $w$ ; i.e. perturb  $w$  by adding random numbers to it.

Advantages: Guarantees that we have a chance of finding global min regardless of starting value of  $w$ .

Disadvantages: If we found the global min we might jump out of it!

In general we cannot tell when we have reached a global min. Hence, we should not turn off noise and stop exploring. Conversely, exploration may be fruitless.