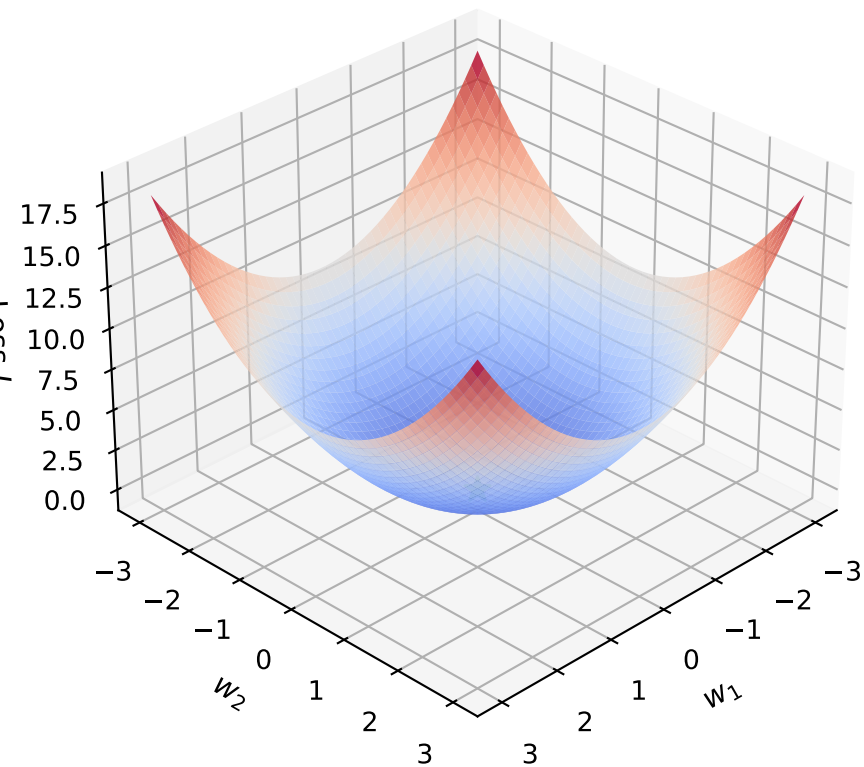


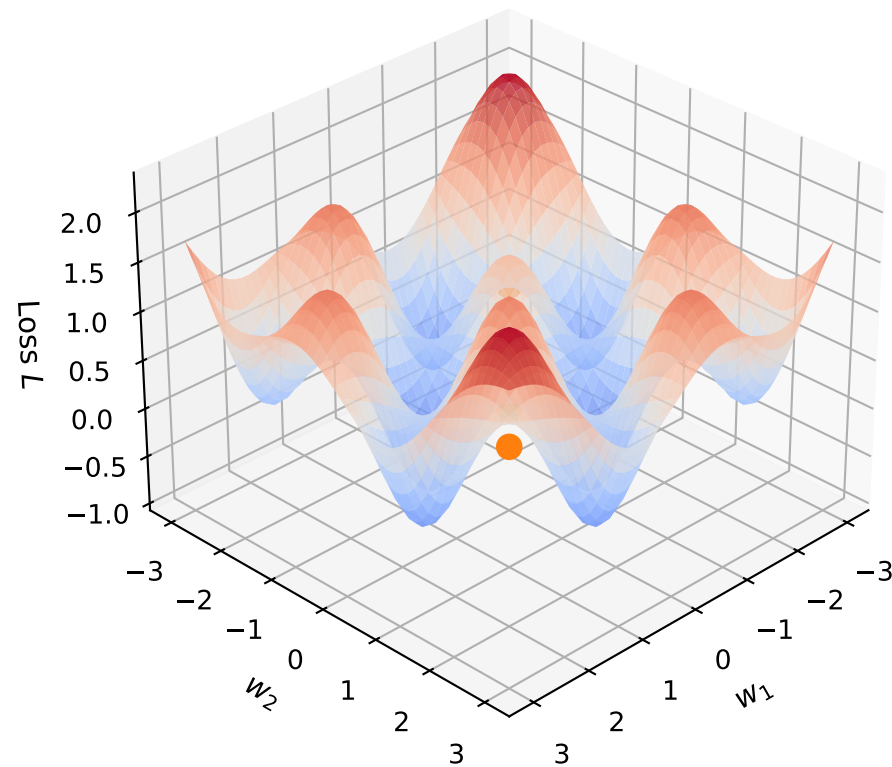
Loss Landscape: Why Deep Networks Are Hard to Train

**Convex Loss Surface
(Single Layer)**



Easy: One global minimum
Gradient descent always finds it

**Non-Convex Loss Surface
(Deep Network)**



Hard: Many local minima
May get stuck in suboptimal solutions