

## Parameter Efficiency: Deep vs. Shallow Networks

### Shallow Network

1000 units

|              |       |
|--------------|-------|
| Layers:      | 1     |
| Total Units: | 1,000 |
| Parameters:  | 1M+   |

### Deep Network

100 units



100 units



100 units

|              |     |
|--------------|-----|
| Layers:      | 3   |
| Total Units: | 300 |
| Parameters:  | 30K |

### Computational Complexity (n inputs)

Shallow

$O(2^n)$

VS

Deep

$O(n^2)$

✓ Better generalization

✓ Faster training

✓ Reduced redundancy

✓ Faster inference