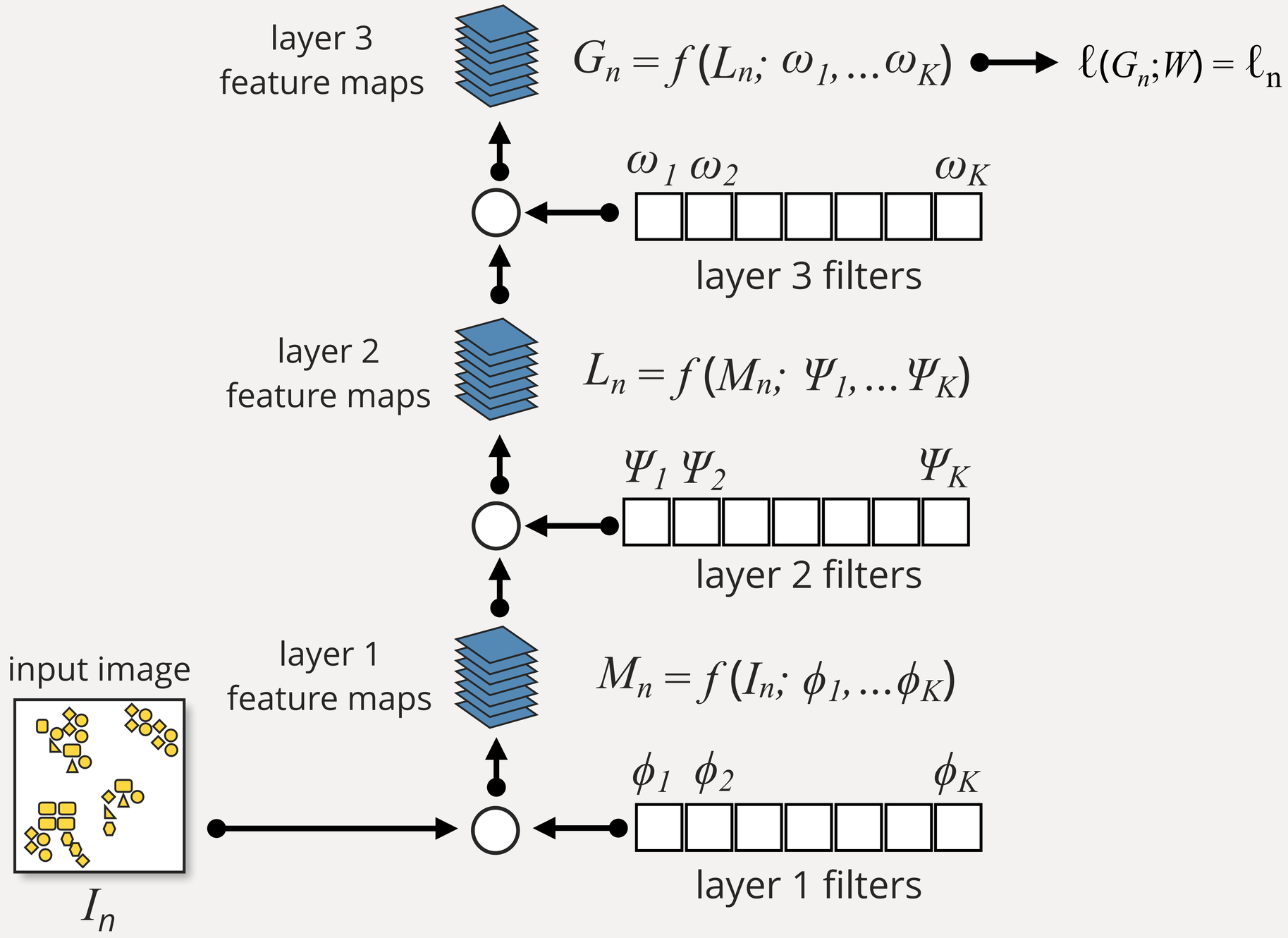
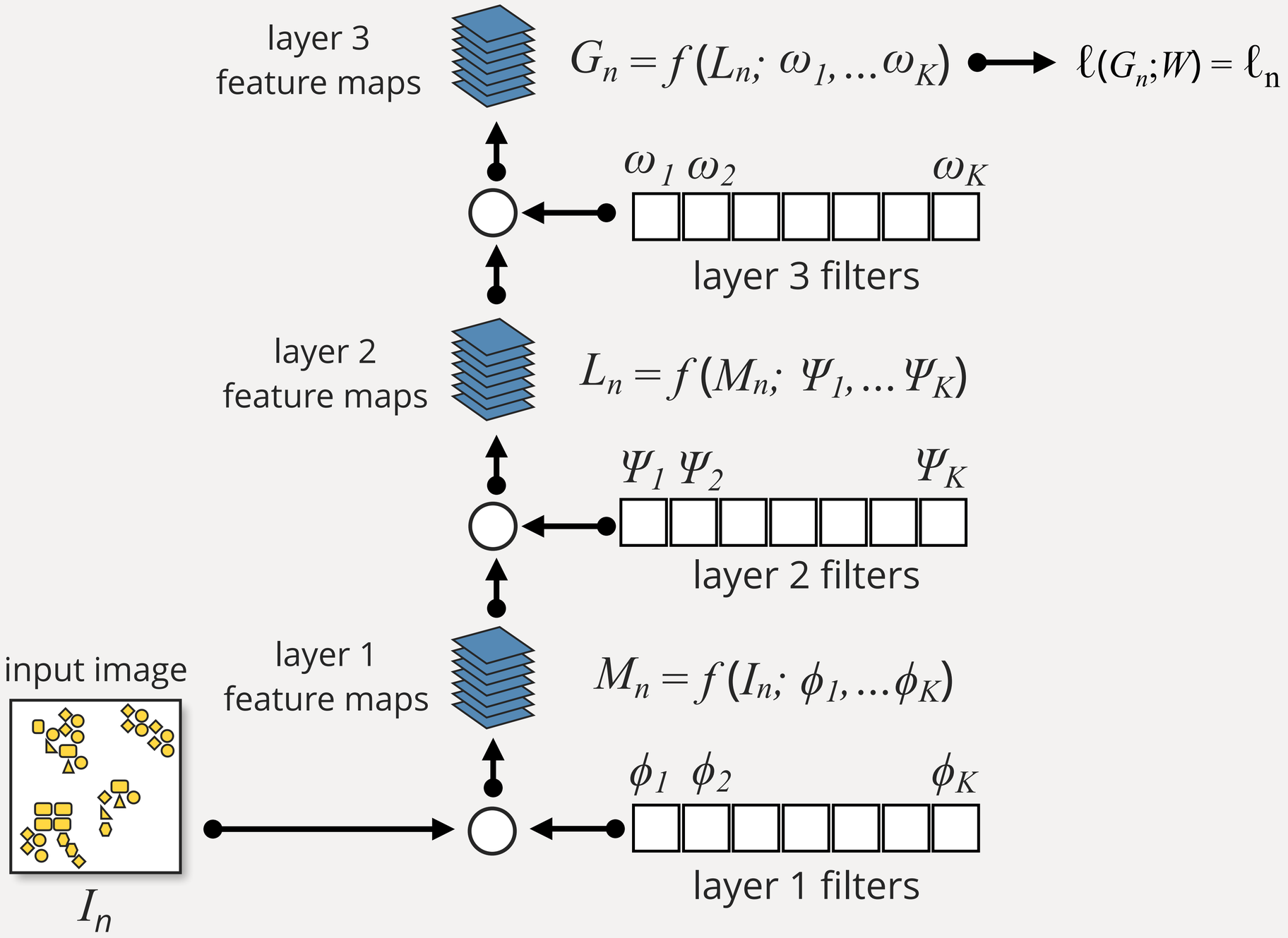
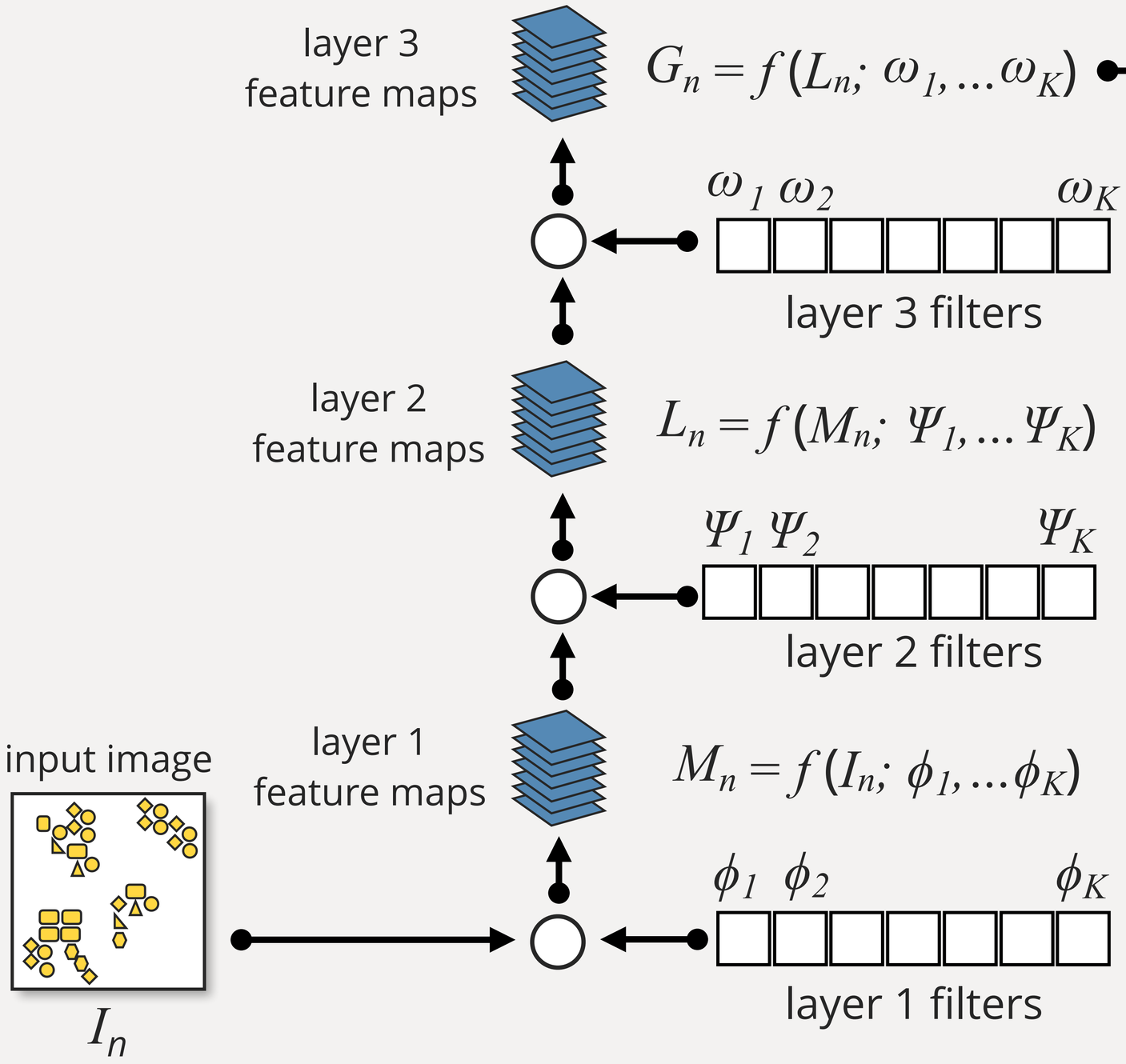
How the Model Learns


layer 3
feature maps
$G_n = f(L_n; \omega_1, \ldots \omega_K)$
$\ell(G_n; W) = \ell_n$
ω_1 ω_2
ω_K
layer 3 filters
layer 2
feature maps
$L_n = f(M_n; \Psi_1, \ldots \Psi_K)$
Ψ_1 Ψ_2
Ψ_K
layer 2 filters
input image
layer 1
feature maps
$M_n = f(I_n; \phi_1, \ldots \phi_K)$
ϕ_1 ϕ_2
ϕ_K
layer 1 filters
I_n


layer 3
feature maps
$G_n = f(L_n; \omega_1, \ldots \omega_K)$
$\ell(G_n; W) = \ell_n$
ω_1 ω_2
ω_K
layer 3 filters
layer 2
feature maps
$L_n = f(M_n; \Psi_1, \ldots \Psi_K)$
Ψ_1 Ψ_2
Ψ_K
layer 2 filters
input image
layer 1
feature maps
$M_n = f(I_n; \phi_1, \ldots \phi_K)$
ϕ_1 ϕ_2
ϕ_K
layer 1 filters
I_n

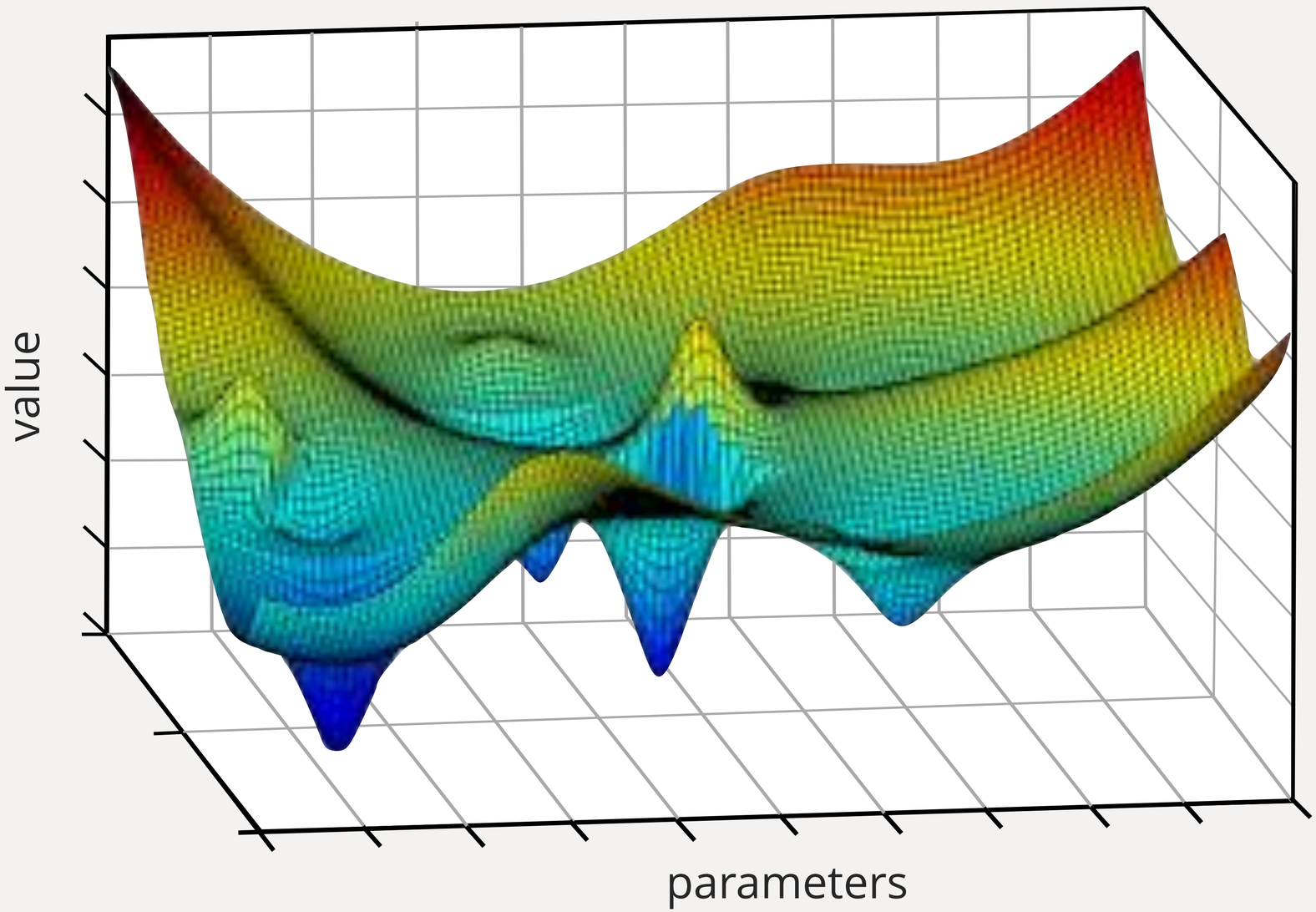


How the Model Learns

- Assume we have labeled data $\{I_n, y_n\}$ $(n=1,N)$
- Assume labels are binary $\in\{+1, -1\}$
- Risk function of model parameters

$$E(\Phi,\Psi,\Omega,W) = 1/\mathrm{N} \sum loss(\mathrm{y_n}, \ell_n)$$

- Find model parameters $\hat{\Phi},\hat{\Psi},\hat{\Omega},\hat{W}$ that minimize $E(\Phi,\Psi,\Omega,W)$


value
parameters