Devising an explicit algorithm based on simple rules is difficult! L1 reg: $\ell = err(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$ favours few non-0 coefs, L2 favours small coefs
under-fitting → high bias (high training, high test error) → add features, decrease regularization term λ, increase degree of polynomial)
over-fitting → high variance (low training, high test error) → get more data, remove features, increase regularization term λ, decrease degree of polynomial)

For convolutional layers:
Image output size = $\lfloor \frac{M+2P-K}{S} \rfloor + 1$
# of params = $C \times K \times K$