1. (b) We are more restricted in our model selection, and thus our hypothesis is less like to fully describe the target function f.

2. (a) $E_{\text{in}} = 0.02857$ and $E_{\text{out}} = 0.084$.

3. (d) $E_{\text{in}} = 0.02857$ and $E_{\text{out}} = 0.08$.

4. (e) $E_{\text{in}} = 0.37$ and $E_{\text{out}} = 0.436$

5. (d) $k = -1$

6. (b) $E_{\text{out}} = 0.056$

7. (c)

8. (d) To calculate the δ's in the second layer, we need

$$\delta_i^{(\ell-1)} = \left(1 - (x_i^{(\ell-1)})^2\right) \sum_{j=1}^{d^\ell} w_{ij}^{(\ell)} \delta_j^{(\ell)}$$

which requires 3 operations. We need to update all the $w_{ij}^{(\ell)}$'s using the rule

$$w_{ij}^{(\ell)} \to w_{ij}^{(\ell)} - \eta \, x_i^{(\ell-1)} \delta_j^{(\ell)}$$

We have a total of $(6 \times 3) + (4 \times 1) = 22$ weights, and thus this requires an additional 22 operations. Afterwards, we update the $x_i^{(\ell)}$'s:

$$x_j^{(\ell)} = \theta \left(\sum_{i=0}^{d^{(\ell-1)}} w_{ij}^{(\ell)} x_i^{(\ell-1)} \right)$$

which again requires 22 operations (number of weights). We get

$$22 + 3 + 22 = 47 \text{ operations.}$$

9. (a) The minimum number is achieved when we have a maximum number of layers with each layer having two nodes. Then the total number of weights is

$$10 + 2 \times (18) = 46.$$

10. (e) To maximize, it should be intuitive that we need 3 layers. Let ℓ be the number of inputs in layer 0. Then we need to maximize

$$10(\ell - 1) + \ell(36 - \ell - 1) + 36 - \ell.$$

Taking the derivative yields

$$10 + 36 - 2\ell - 2 = 0 \implies \ell = 22$$

so the maximum number of weights is 510.