



deeplearning.ai

Regularizing your  
neural network

---

**Regularization**

# Logistic regression

$$\min_{w,b} J(w,b)$$

$$\underline{w \in \mathbb{R}^{n_x}}, \underline{b \in \mathbb{R}}$$

$\lambda =$  regularization parameter  
lambda lambda

$$J(w,b) = \underbrace{\frac{1}{m} \sum_{i=1}^m \ell(y^{(i)}, \hat{y}^{(i)})}_{\text{cost function}} + \frac{\lambda}{2m} \underbrace{\|w\|_2^2}_{\text{L2 regularization}}$$

~~$$+ \frac{\lambda}{2m} b^2$$~~  
 omit

$L_2$  regularization  $\underline{\|w\|_2^2} = \sum_{j=1}^{n_x} w_j^2 = w^T w \leftarrow$

$L_1$  regularization  $\frac{\lambda}{2m} \sum_{j=1}^{n_x} |w_j| = \frac{\lambda}{2m} \|w\|_1$

$w$  will be sparse

# Neural network

$$\rightarrow J(w^{[1]}, b^{[1]}, \dots, w^{[L]}, b^{[L]}) = \underbrace{\frac{1}{m} \sum_{i=1}^m \ell(y^{(i)}, \hat{y}^{(i)})}_{\text{loss}} + \underbrace{\frac{\lambda}{2m} \sum_{l=1}^L \|w^{[l]}\|_F^2}_{\text{weight decay}}$$

$$\|w^{[l]}\|_F^2 = \sum_{i=1}^{n^{[l]}} \sum_{j=1}^{n^{[l-1]}} (w_{ij}^{[l]})^2$$

$w^{[l]}: \begin{matrix} n^{[l]} & n^{[l-1]} \\ \uparrow & \uparrow \end{matrix}$

"Frobenius norm"

$\|\cdot\|_2^2$

$\|\cdot\|_F^2$

$$dw^{[l]} = \left[ \text{(from backprop)} + \frac{\lambda}{m} w^{[l]} \right]$$

$$\rightarrow w^{[l]} := w^{[l]} - \alpha \underbrace{dw^{[l]}}_{\text{weight decay}}$$

$$\underbrace{\frac{\partial J}{\partial w^{[l]}}}_{\text{weight decay}} = dw^{[l]}$$

"Weight decay"

$$w^{[l]} := w^{[l]} - \alpha \left[ \text{(from backprop)} + \frac{\lambda}{m} w^{[l]} \right]$$

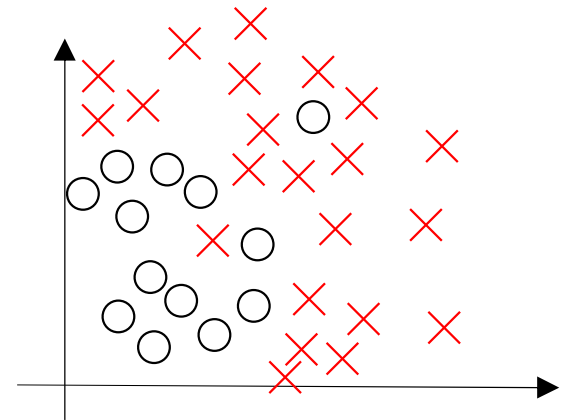
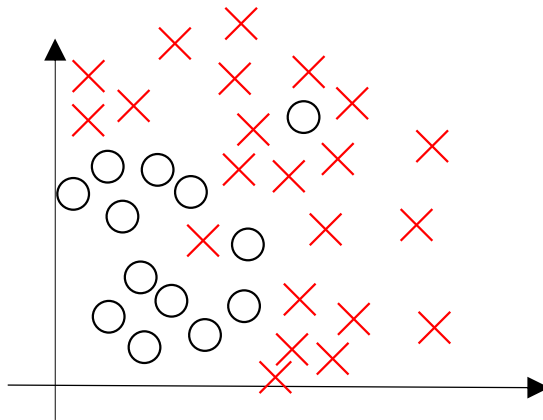
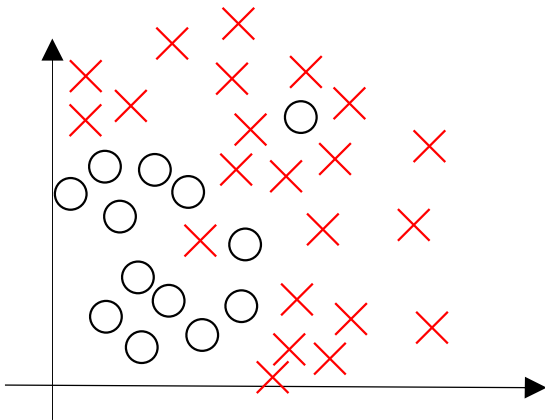
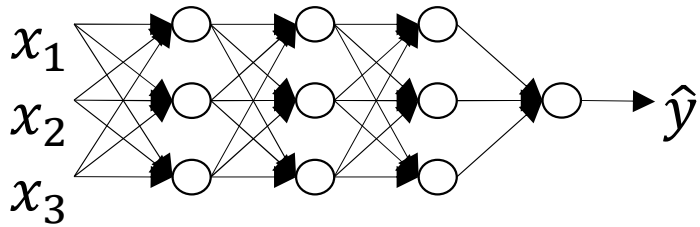
$$= w^{[l]} - \frac{\alpha \lambda}{m} w^{[l]} - \alpha \text{(from backprop)}$$

$$= \underbrace{\left(1 - \frac{\alpha \lambda}{m}\right)}_{< 1} \underbrace{w^{[l]}}_{\text{weight decay}} - \alpha \text{(from backprop)}$$

# Neural network

$$J(w^{[1]}, b^{[1]}, \dots, w^{[L]}, b^{[L]}) = \frac{1}{n} \sum_{i=1}^n \ell(y^{(i)}, \hat{y}^{(i)}) + \frac{\lambda}{2n} \sum_{l=1}^L \|w^{[l]}\|^2$$

# How does regularization prevent overfitting?



# How does regularization prevent overfitting?