

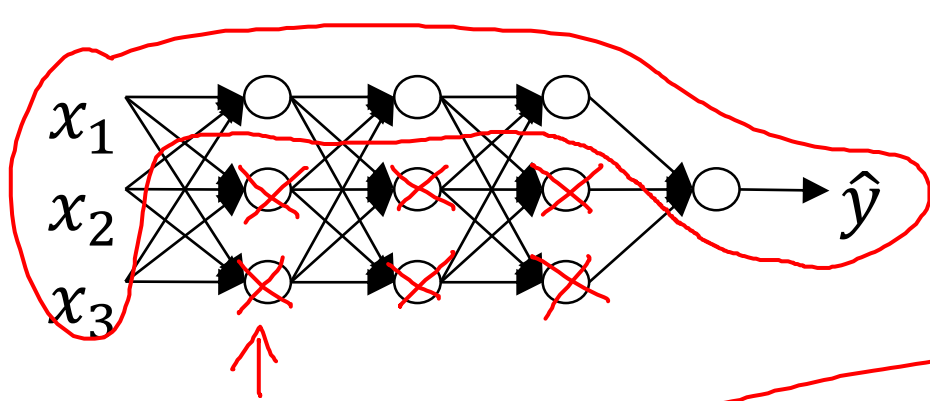


deeplearning.ai

Regularizing your neural network

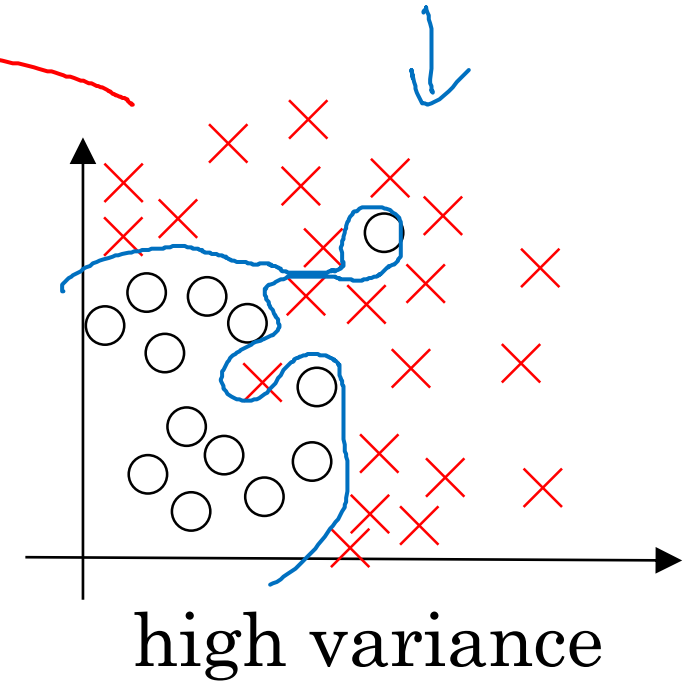
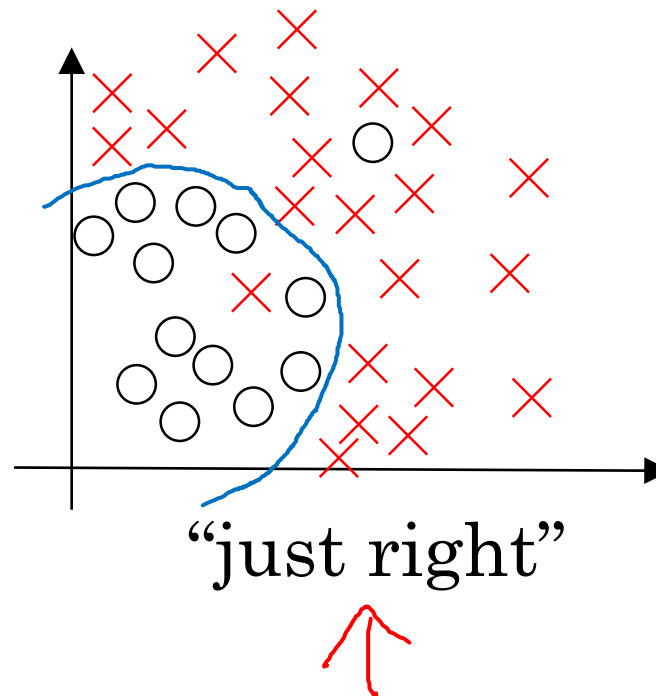
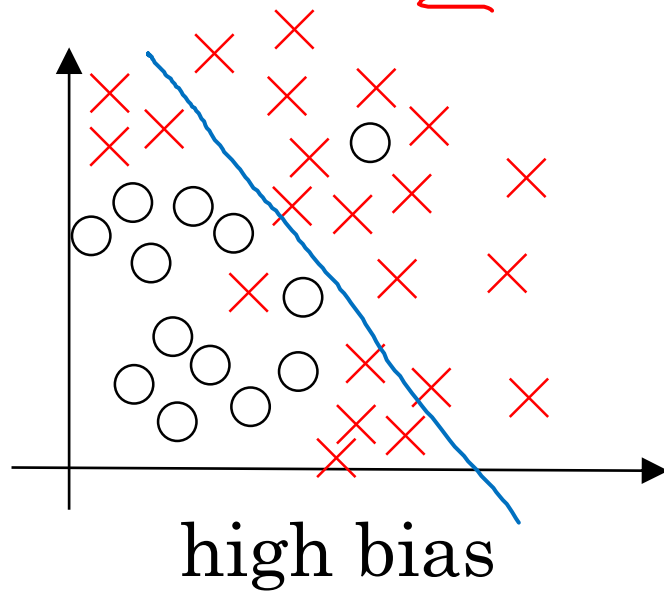
Why regularization reduces overfitting

How does regularization prevent overfitting?

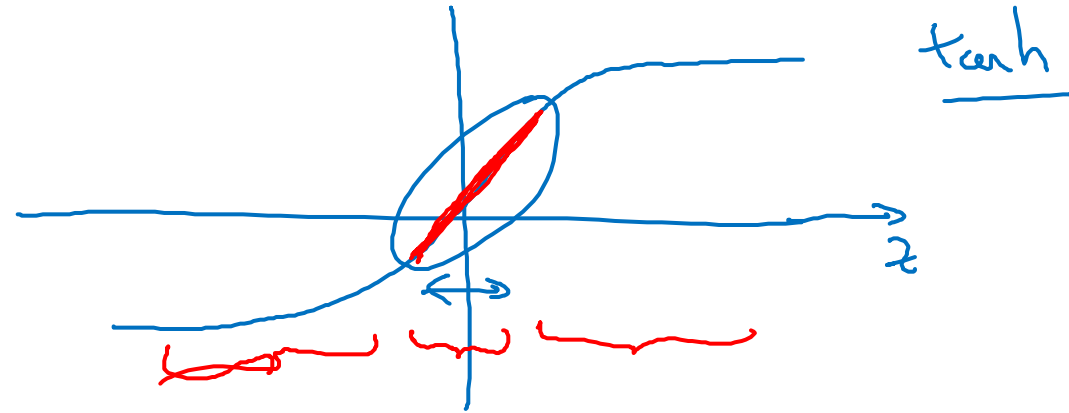


$$J(\mathbf{w}^{(1)}, \mathbf{b}^{(1)}) = \frac{1}{n} \sum_{i=1}^n \ell(y^{(i)}, \hat{y}^{(i)}) + \frac{\lambda}{2n} \sum_{l=1}^L \underbrace{\|\mathbf{w}^{(l)}\|_F^2}_{\text{regularization}}$$

$$\mathbf{w}^{(1)} \approx 0$$



How does regularization prevent overfitting?



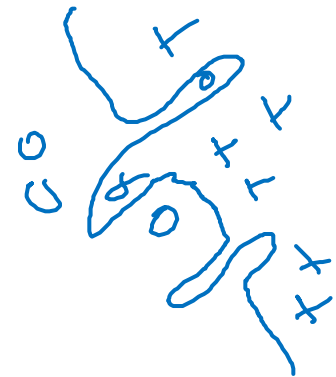
$$g(z) = \tanh(z)$$

$\lambda \uparrow$

$W^{[L]} \downarrow$

$$z^{[L]} = \underline{W}^{[L]} a^{[L-1]} + \underline{b}^{[L]}$$

Every layer \approx linear.



$$J(\dots) = \underbrace{\sum_i \mathcal{L}(\hat{y}^{(i)}, y^{(i)})}_{\text{training loss}} + \underbrace{\frac{\lambda}{2m} \sum_L \|W^{[L]}\|_F^2}_{\text{regularization term}}$$

