Regularizing your neural network

Why regularization reduces

How does regularization prevent overfitting?

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

$$w^{[l]} \approx 0$$

high bias   "just right"   high variance

How does regularization prevent overfitting?

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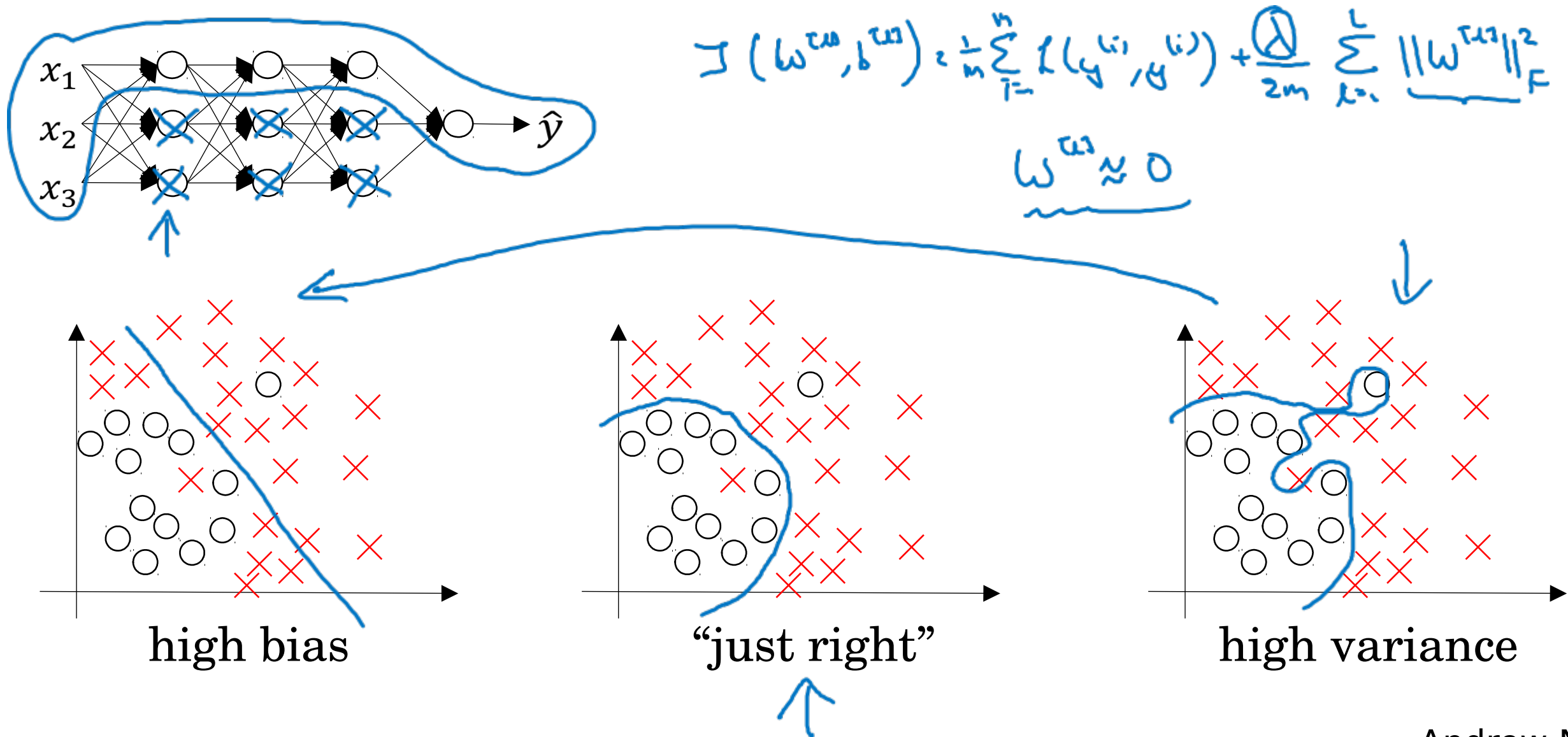
$g(z) = \tanh(z)$

$\lambda \uparrow$ $W^{[l]} \downarrow$ $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$

Every layer $\approx$ linear.

$$J(\cdots) = \boxed{\sum_i \mathcal{L}(\hat{y}^{(i)}, y^{(i)})} \; \frac{\lambda}{2m} \sum_l \|W^{[l]}\|_F^2$$

Andrew Ng