



3

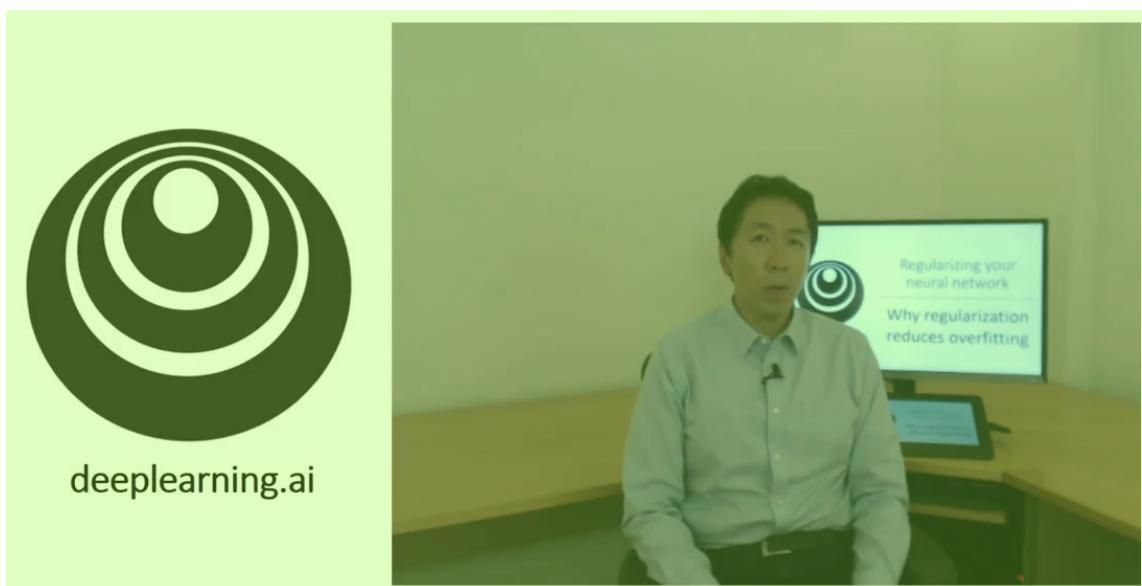
Regularizing your<sup>2</sup>  
neural network

---

Why regularization<sup>4</sup>  
reduces overfitting

↳

[■■■: 00:00:03.000 1



[■ 2 ■ / ■ 26 ■  
■■: keyframe\_00-00-03-000\_0001.jpg



3

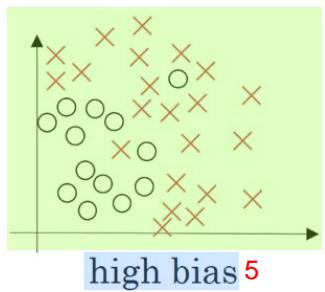
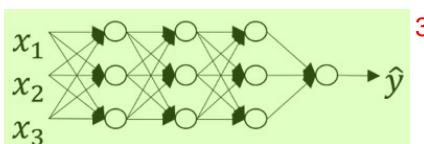
Regularizing your<sup>2</sup>  
neural network

---

Why regularization<sup>4</sup>  
reduces overfitting

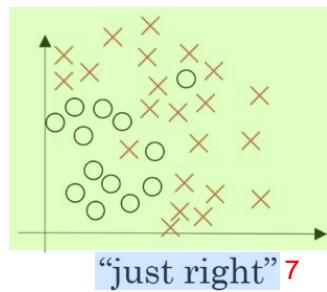
↳

## How does regularization prevent overfitting?<sup>2</sup>



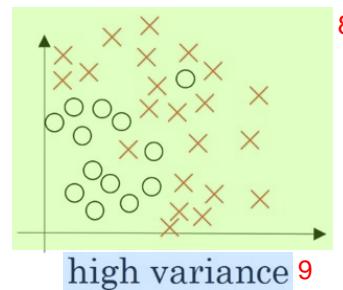
4

high bias 5



6

“just right” 7

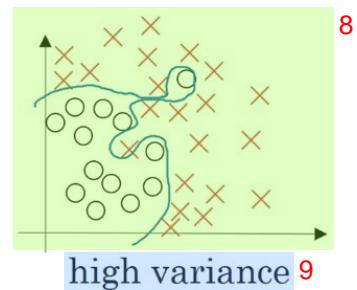
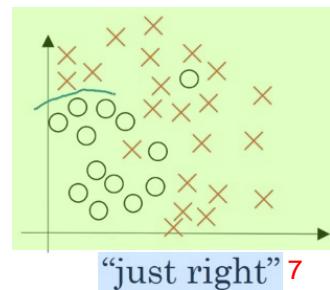
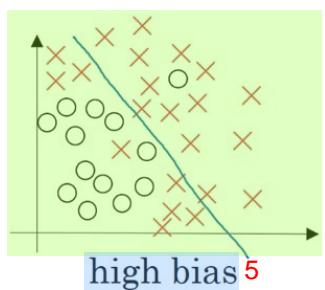
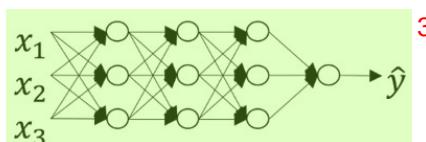


8

high variance 9

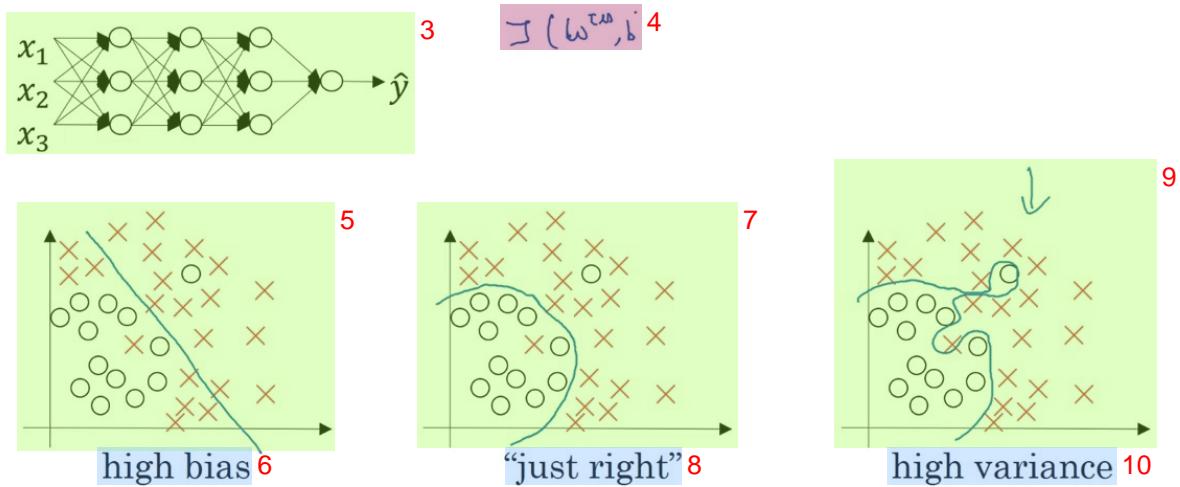
Andrew Ng 10

## How does regularization prevent overfitting?<sup>2</sup>



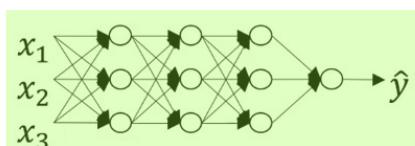
Andrew Ng 10

## How does regularization prevent overfitting? <sup>2</sup>



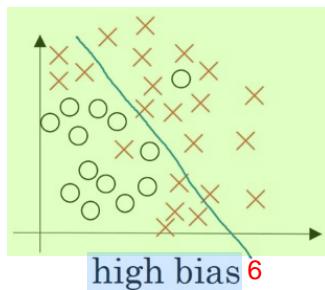
Andrew Ng 11

## How does regularization prevent overfitting?<sup>2</sup>

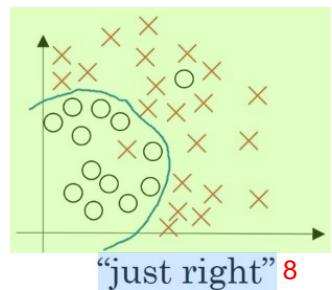


3

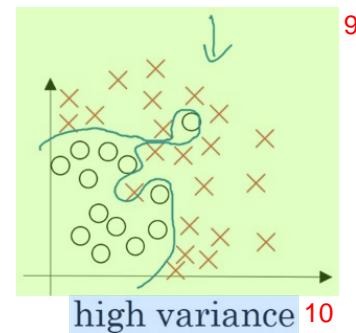
$$J(\omega^{(0)}, b^{(0)}) = \frac{1}{m} \sum_{i=1}^m L(y^{(i)}, \hat{y}^{(i)}) + \frac{\lambda}{2m} \sum_{l=1}^L \|\omega^{(l)}\|^2$$



5



7

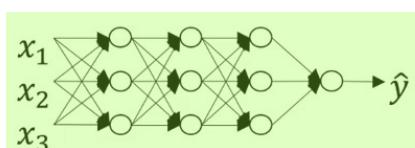


9

high variance 10

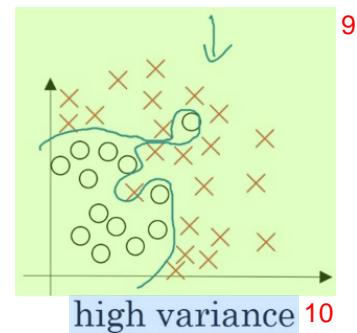
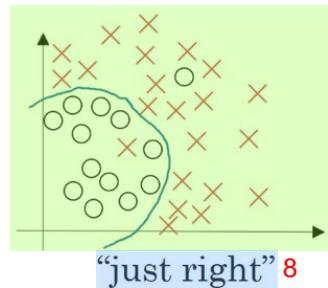
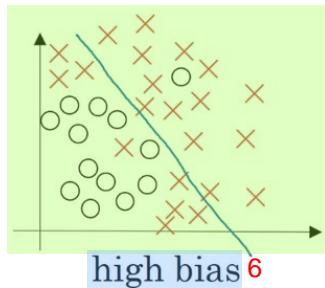
Andrew Ng 11

## How does regularization prevent overfitting? <sup>2</sup>



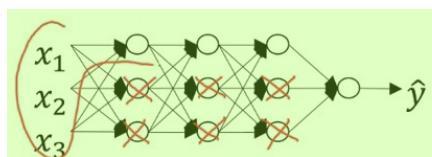
$$J(\omega^{(m)}, b^{(m)}) = \frac{1}{m} \sum_{i=1}^m L(y_i^{(i)}, \hat{y}_i^{(i)}) + \frac{\lambda}{2m} \sum_{l=1}^L \|\omega^{(l)}\|_F^2$$

4



Andrew Ng 11

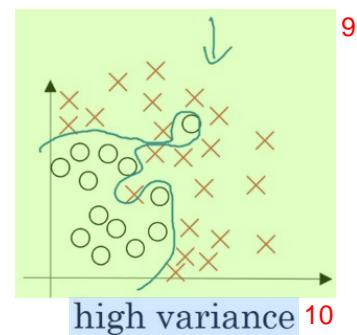
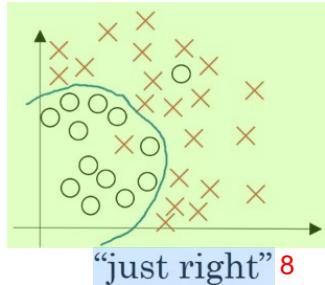
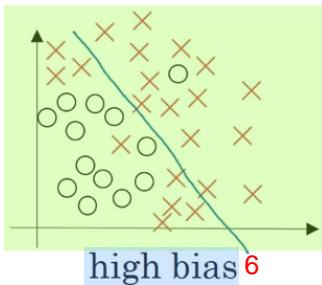
## How does regularization prevent overfitting? <sup>2</sup>



$$J(\omega^{(m)}, b^{(m)}) = \frac{1}{m} \sum_{i=1}^m L(y_i^{(i)}, \hat{y}_i^{(i)}) + \frac{\lambda}{2m} \sum_{l=1}^L \|\omega^{(l)}\|_F^2$$

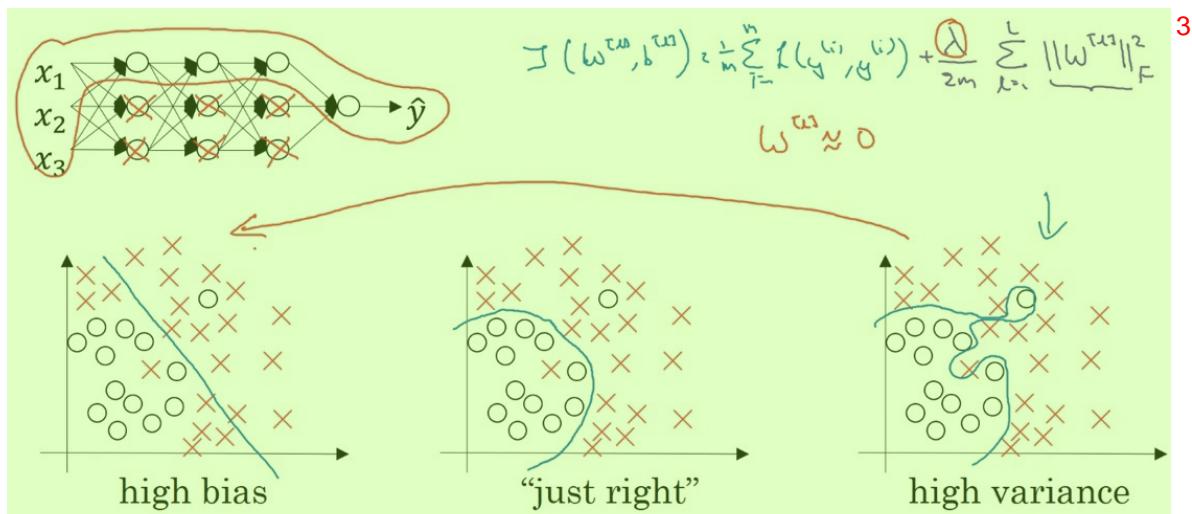
$\omega^{(l)} \approx 0$

4



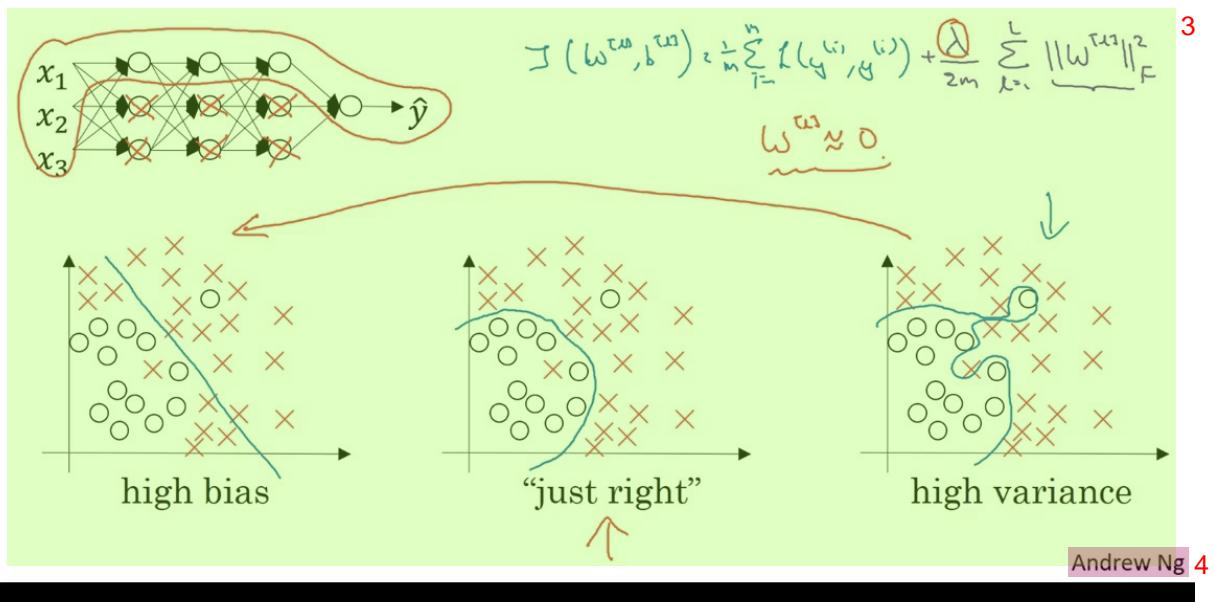
Andrew Ng 11

## How does regularization prevent overfitting?<sup>2</sup>

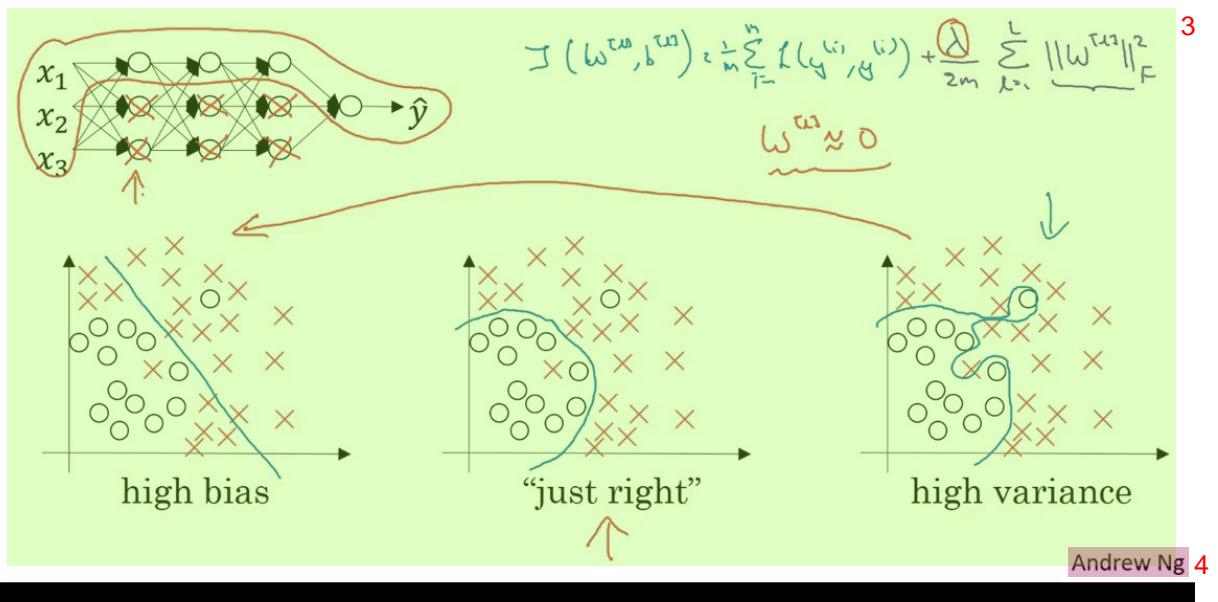


Andrew Ng 4

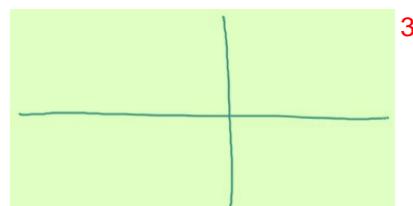
## How does regularization prevent overfitting? <sup>2</sup>



## How does regularization prevent overfitting? <sup>2</sup>



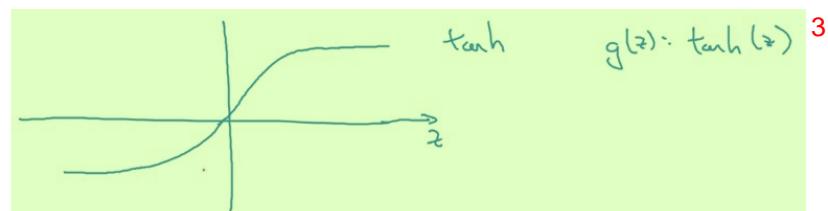
How does regularization prevent overfitting?<sup>2</sup>



3

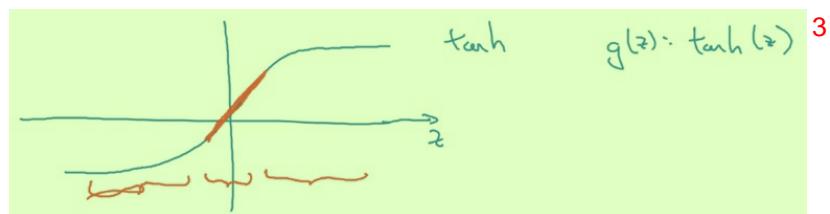
Andrew Ng

## How does regularization prevent overfitting?<sup>2</sup>



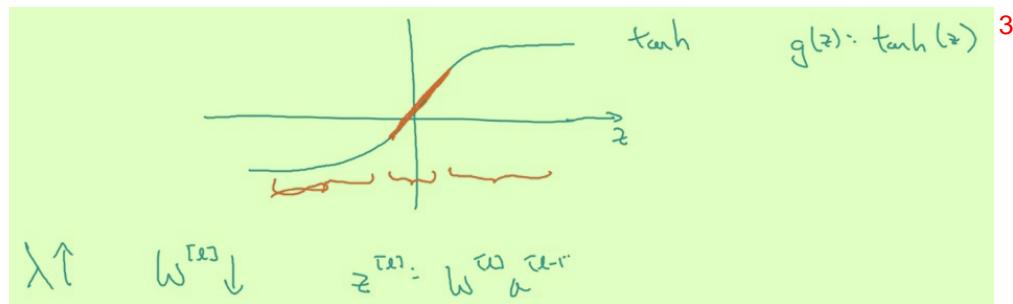
Andrew Ng

## How does regularization prevent overfitting?<sup>2</sup>



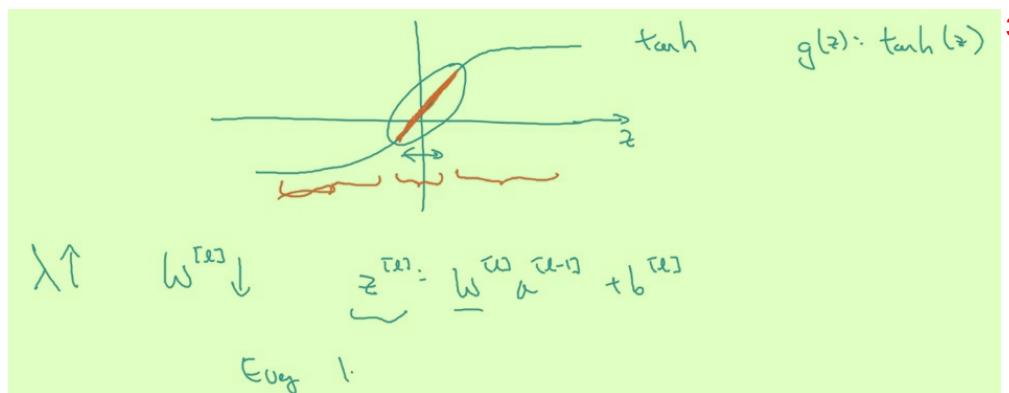
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



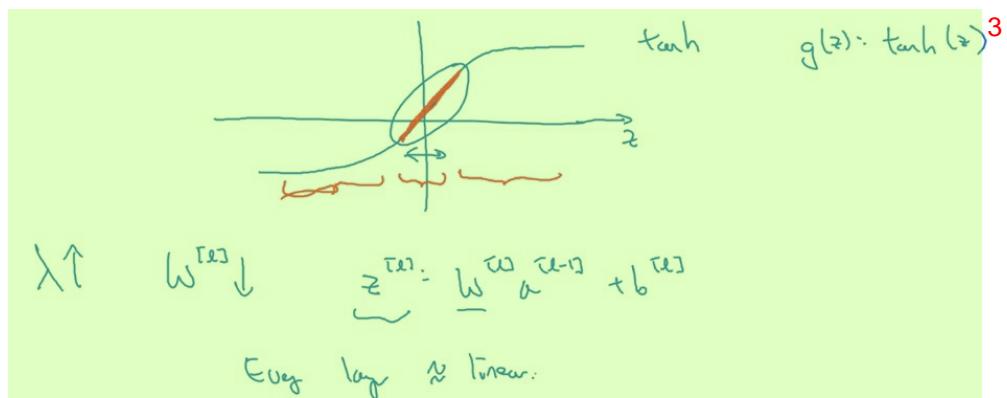
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



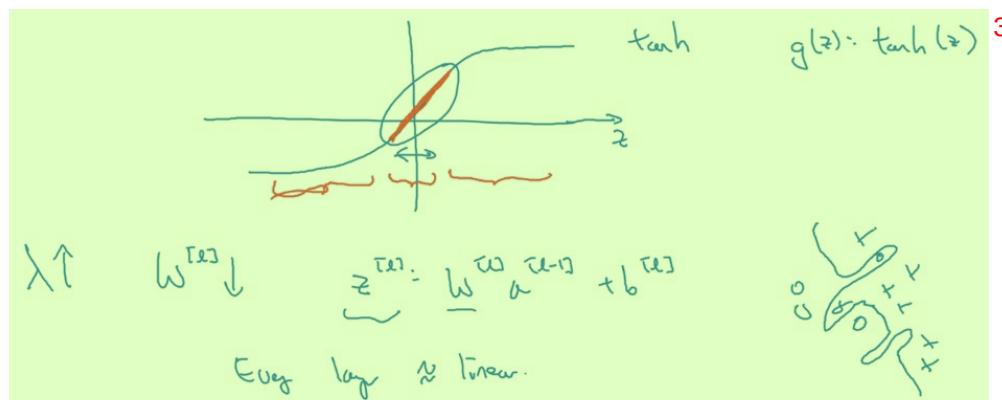
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



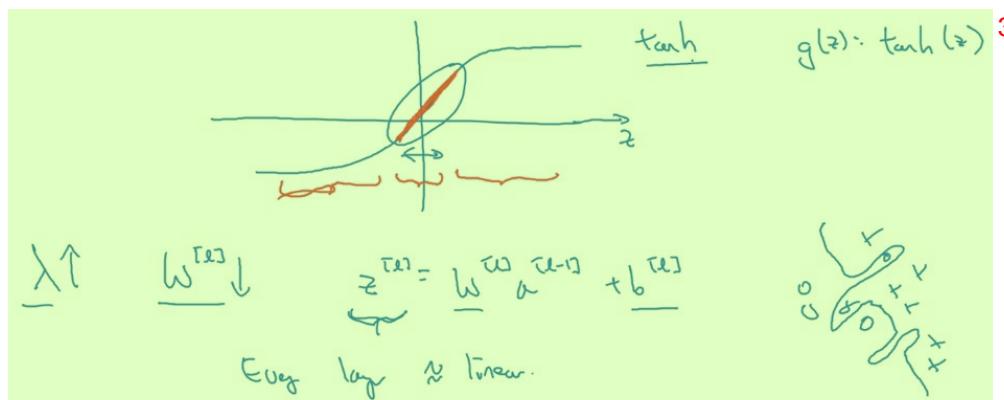
Andrew Ng

## How does regularization prevent overfitting?<sup>2</sup>



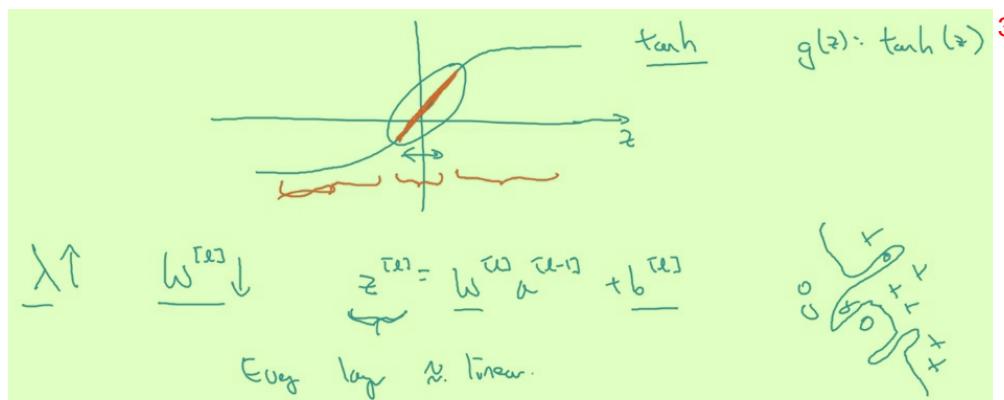
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



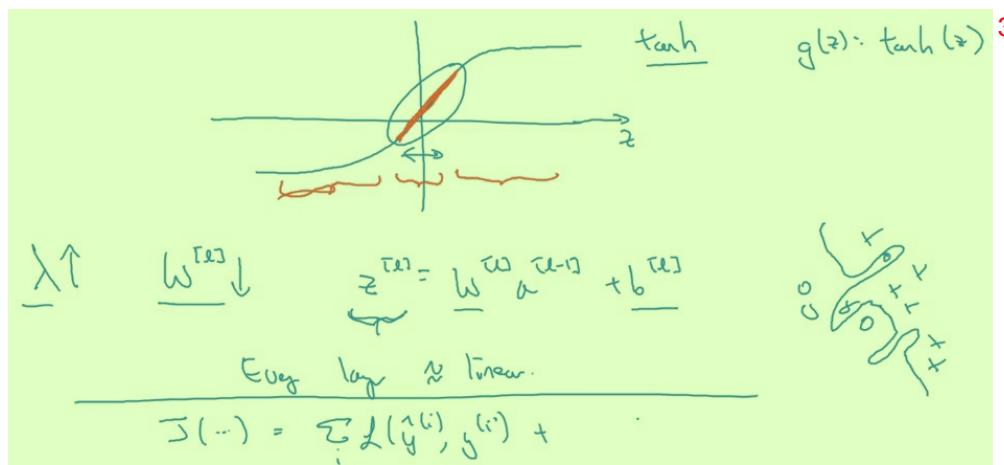
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



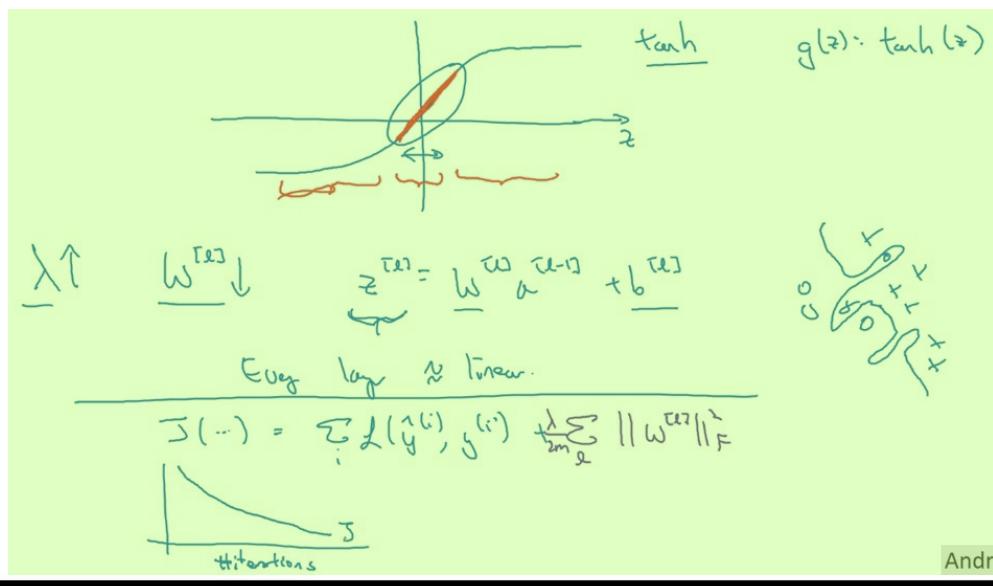
Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>

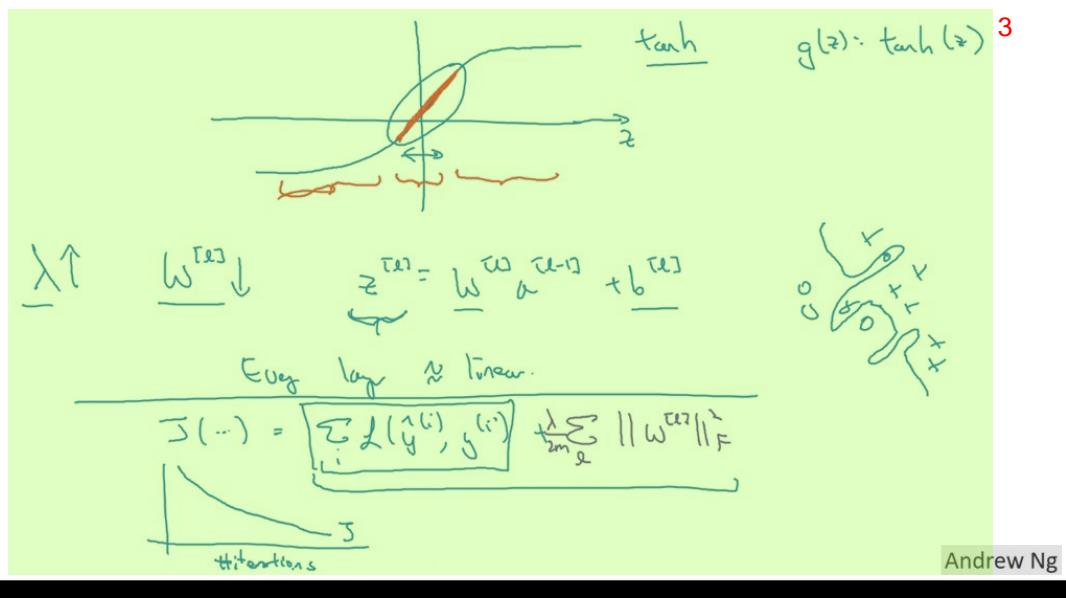


Andrew Ng

## How does regularization prevent overfitting? <sup>2</sup>



## How does regularization prevent overfitting? <sup>2</sup>



[■■■: 00:06:54.000] 1

How does regularization prevent overfitting? 2 3 4 5



[■■■: 00:07:00.000] 1

