

Choose model

2023年5月24日 20:52

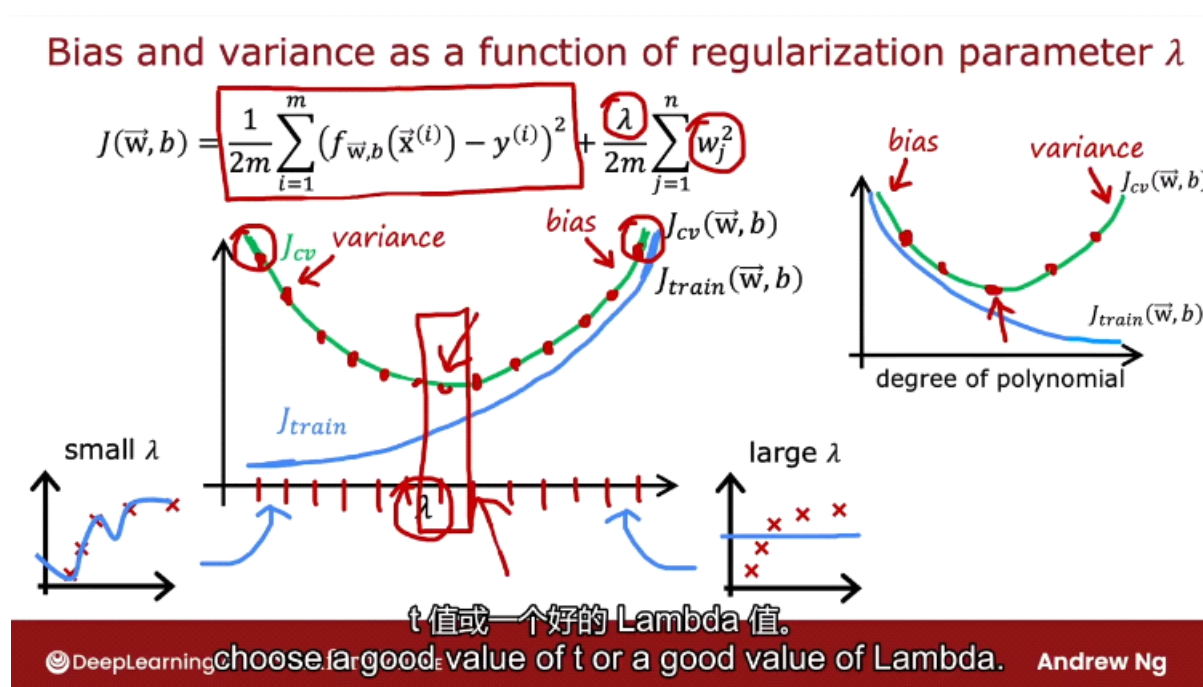
1.选择交叉验证数据集，用于选择合适的模型，选出合适模型后再利用新的验证集去评估模型的性能

(交叉验证相当于对不同模型选择的一次学习，倾向于过拟合，不能用来评估模型的泛化程度)

2.模型的Bias与Variance，来判断是欠拟合还是过拟合

(Both High，可能是部分过拟合而部分欠拟合，很糟糕)

3.正则化时，尝试不同 λ ，用交叉验证集选择合适的 λ ，再用test set评估。



4.建立性能评估的基准:

Establishing a baseline level of performance

What is the level of error you can reasonably hope to get to?

- Human level performance
- Competing algorithms performance
- Guess based on experience

Bias/variance examples

Baseline performance	: 10.6%				
Training error (J_{train})	: 10.8%	0.2%	10.6%	4.4%	10.6%
Cross validation error (J_{cv})	: 14.8%	4.0%	15.0%	0.5%	15.0%
			15.5%		19.7%
		high variance	high bias		high bias high variance

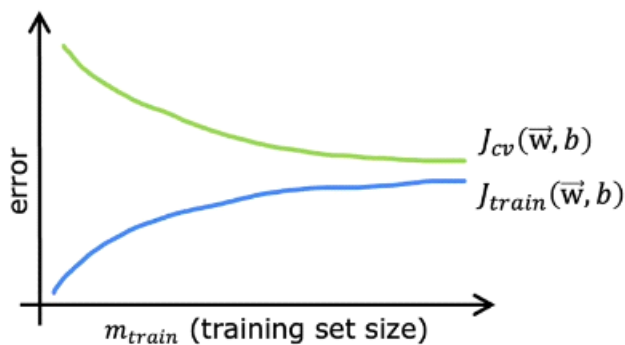
5.学习曲线

误差随训练集大小变化的曲线

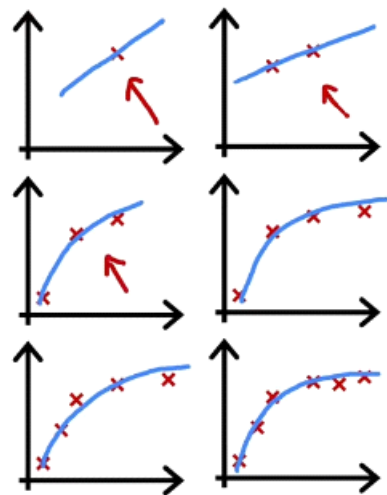
Learning curves

J_{train} = training error

J_{cv} = cross validation error



$$f_{\bar{w},b}(x) = w_1x + w_2x^2 + b$$



或非常小的训练误差,

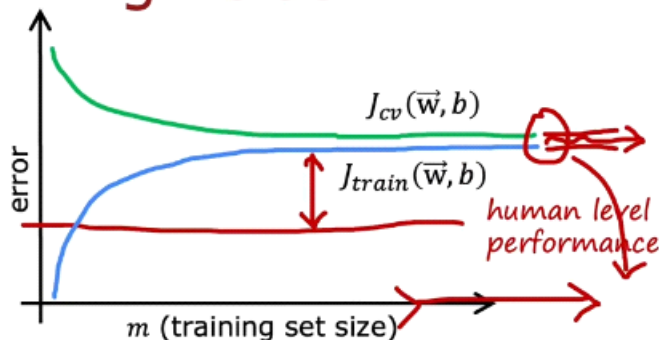
DeepLearning.AI

Stanford

get zero or very small training error,

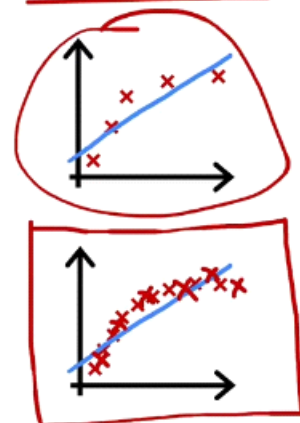
Andrew Ng

High bias

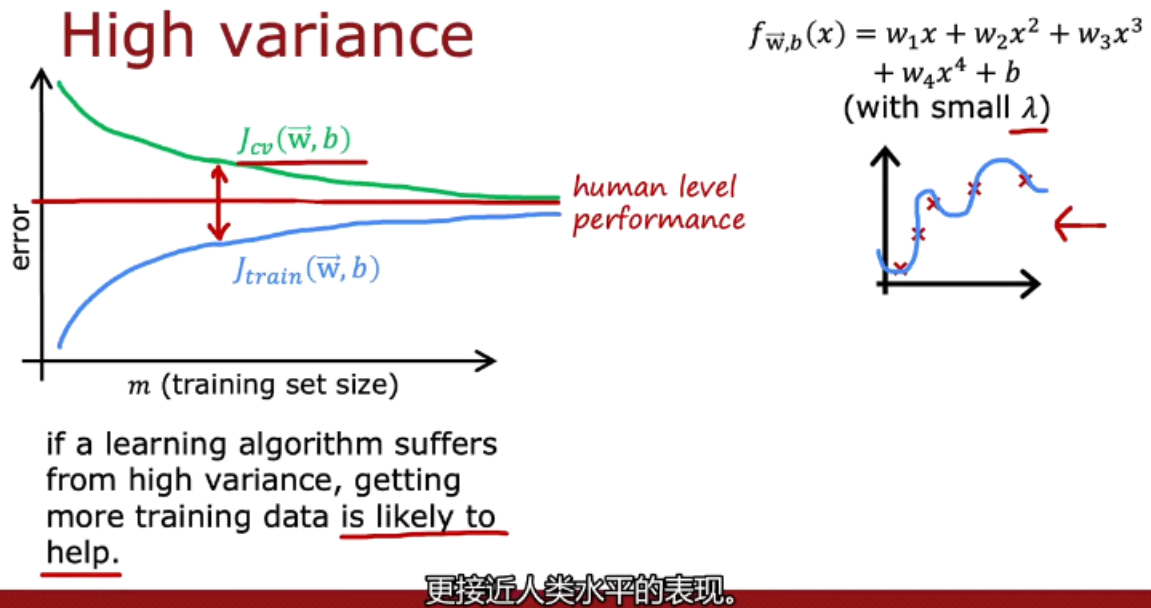


if a learning algorithm suffers from high bias, getting more training data will not (by itself) help much.

$$f_{\bar{w},b}(x) = w_1x + b$$



High Bias情况，数据集再大都不会有用



High Variance情况，数据集变大会让模型变更好

6. 对一个线性回归模型，不同解决方法解决的问题：

Debugging a learning algorithm

You've implemented regularized linear regression on housing prices

$$J(\vec{w}, b) = \frac{1}{2m} \sum_{i=1}^m (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2m} \sum_{j=1}^n w_j^2$$

But it makes unacceptably large errors in predictions. What do you try next?

- Get more training examples
- Try smaller sets of features x, x^2, x^3, x^4, \dots
- Try getting additional features
- Try adding polynomial features $(x_1^2, x_2^2, x_1x_2, \text{etc})$
- Try decreasing λ
- Try increasing λ

fixes high variance
fixes high variance
fixes high bias
fixes high bias
fixes high bias
fixes high variance

屏幕剪辑的捕获时间: 2023/5/25 8:50

High Bias情况应该给予模型更高的灵活性

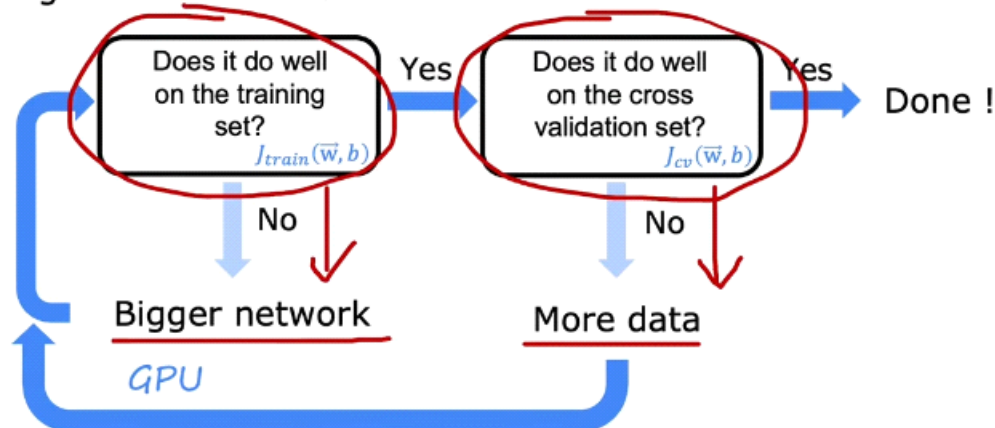
High Variance情况应该减小模型的灵活性，使其变得更平滑

7. Bias and Variance在神经网络中的应用

前提：如果神经网络足够大，则一定是Low Bias模型

Neural networks and bias variance

Large neural networks are low bias machines



Neural network regularization

$$J(\mathbf{W}, \mathbf{B}) = \frac{1}{m} \sum_{i=1}^m L(f(\vec{x}^{(i)}), y^{(i)}) + \frac{\lambda}{2m} \sum_{\text{all weights } \mathbf{W}} (w^2)$$

Unregularized MNIST model

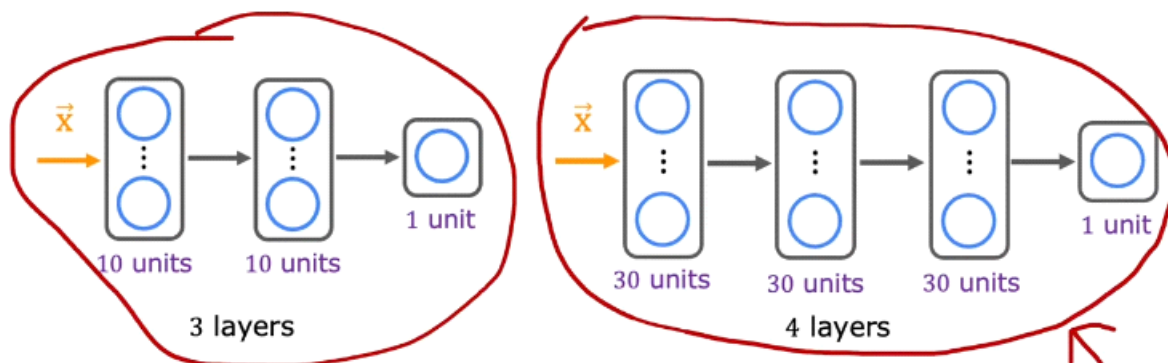
```
layer_1 = Dense(units=25, activation="relu")
layer_2 = Dense(units=15, activation="relu")
layer_3 = Dense(units=1, activation="sigmoid")
model = Sequential([layer_1, layer_2, layer_3])
```

Regularized MNIST model

```
layer_1 = Dense(units=25, activation="relu", kernel_regularizer=L2(0.01))
layer_2 = Dense(units=15, activation="relu", kernel_regularizer=L2(0.01))
layer_3 = Dense(units=1, activation="sigmoid", kernel_regularizer=L2(0.01))
model = Sequential([layer_1, layer_2, layer_3])
```

神经网络正则化（用得并不多）

Neural networks and regularization



A large neural network will usually do as well or better than a smaller one so long as regularization is chosen appropriately.

用更大的（正则化的）网络几乎永远没有坏处，除了会让计算更慢