Neural network is very famous because it has new ways to address high bias/variance.

## The bias variance tradeoff

There is a tradeoff between a simple model and a complex model.

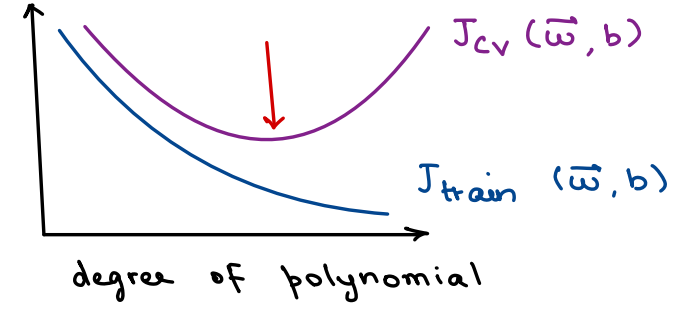
$f_{\vec{w},b}(\vec{x}) = w_1 x + b$

simple model

$f_{\vec{w},b}(x) = w_1 x + w_2 x^2 + b$

$f_{\vec{w},b}(\vec{x}) = w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$

complex model

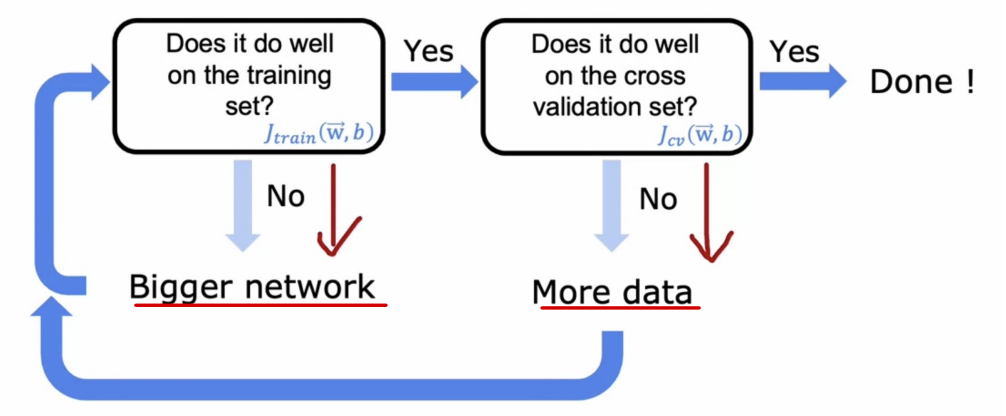



With a simple model we have a high bias problem and a complex model we have a high variance.

A neural network has a simple recipe for solving this.

First of all a large enough neural network will always have a low bias unless the size of the training data is enormous.

Recipe:

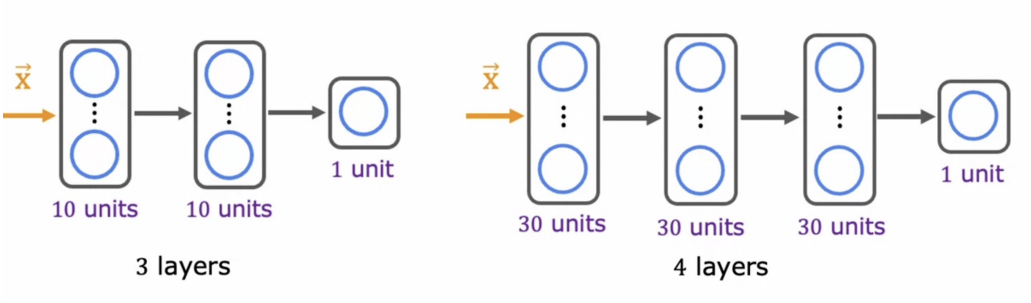



Limitations:

Bigger network: A bigger network means that cost of computation will increase making it very expensive. That is why rise of neural networks is often linked to rise of GPUs.

More data: It is not possible to get a lot of data beyond a certain point.

Does a very large neural network result in a high variance problem?




If regularization is chosen appropriately then in almost all cases a large neural network will do as good or better than a smaller one.

# The only problem with having a large neural network its computationally more expensive and slows down the algorithm.

To regularize a neural network in tenserflow in code:

## Neural network regularization

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

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])
```

value for $\lambda$ can be chosen differently