

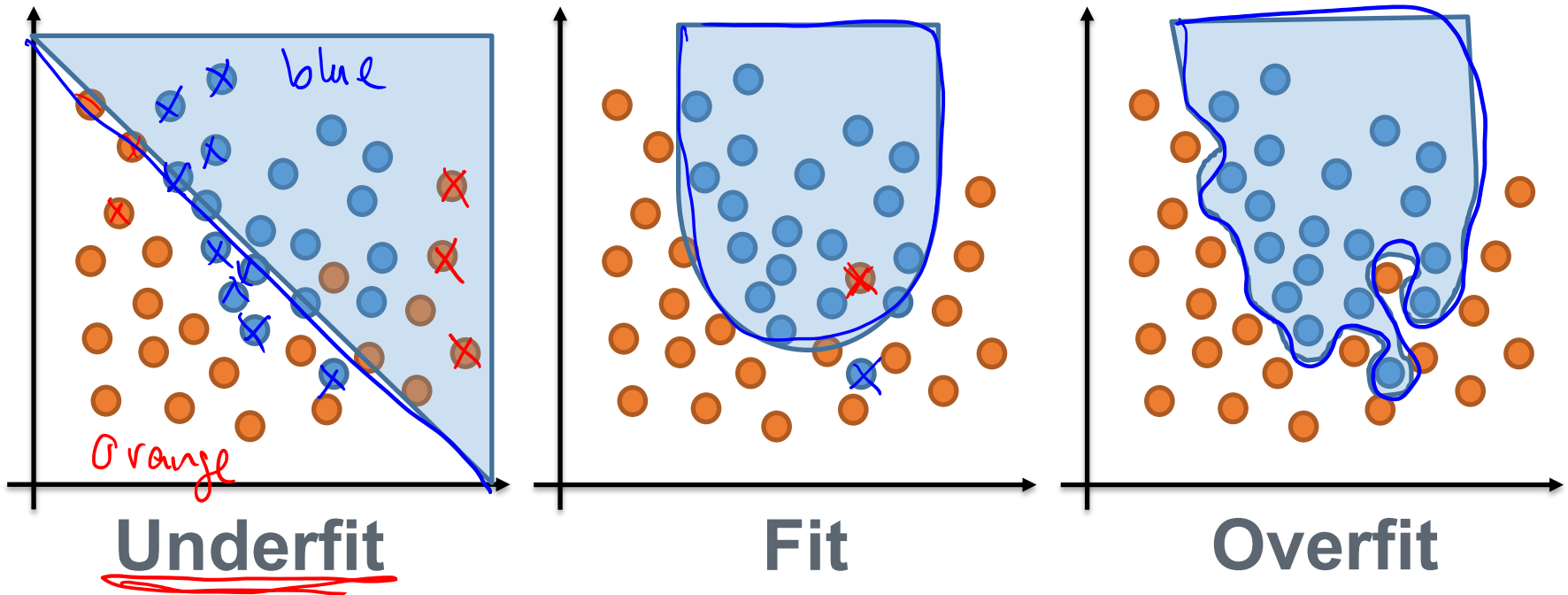


Dropout

Heni Ben Amor, Ph.D.
Assistant Professor
Arizona State University

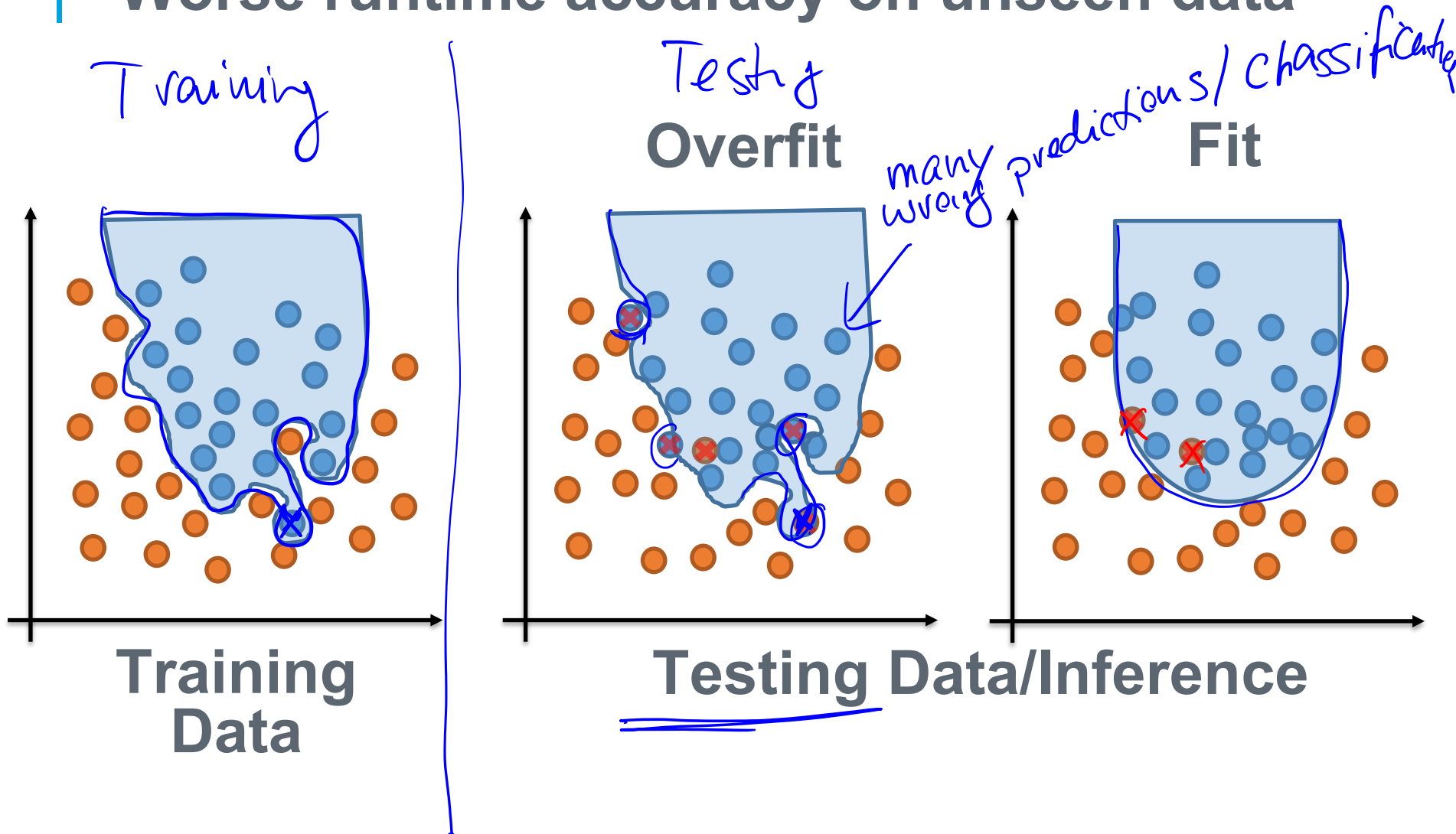
Overfitting

- Noise in training data
- Network can learn to model this noise



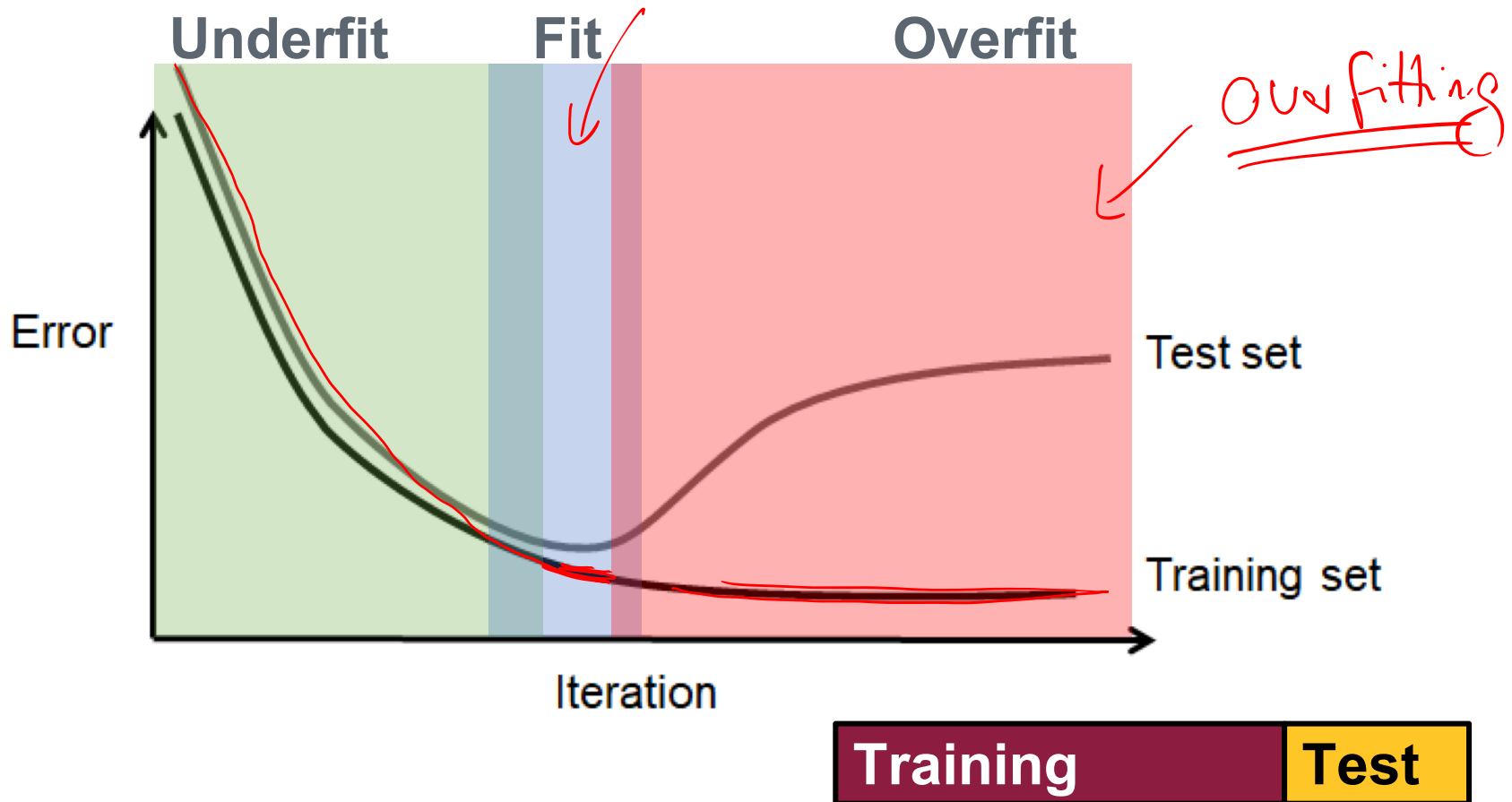
Overfitting

Worse runtime accuracy on unseen data



Solutions: Early stopping

Stop training when loss on testing dataset begins increasing.



Solutions: Dropout

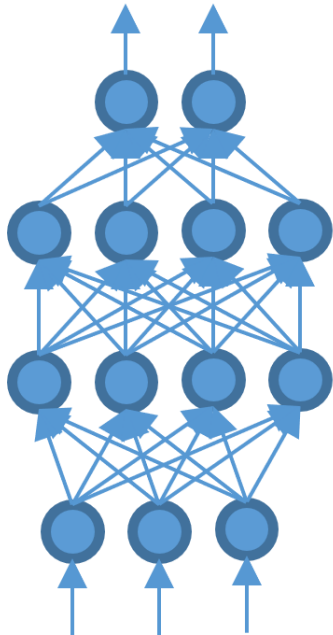


- | Introduced by Geoffrey Hinton et al. in 2012
- | Mean of many differently trained networks likelier to produce better results.
- | Dropout approximates using multiple neural networks without the costly computation or memory.

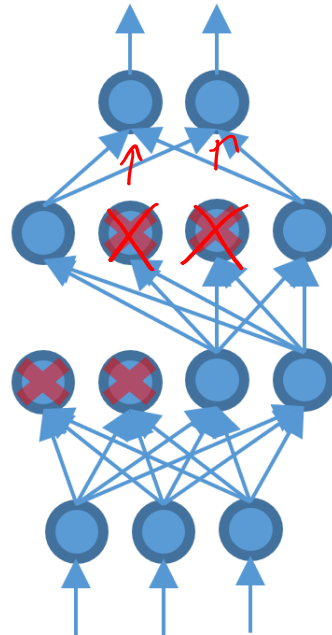
Dropout Training

During training, for any number of layers, completely dropout each neuron with probability P .

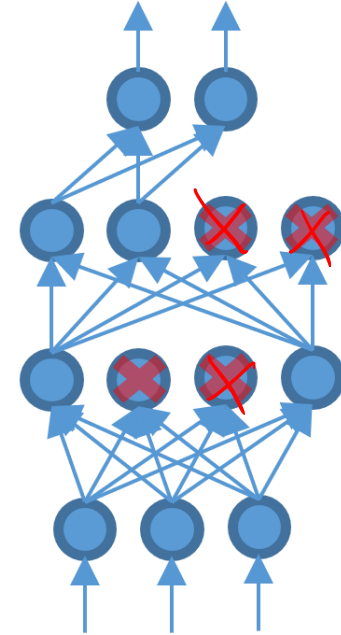
Standard Fully Connected



Using dropout with $P = .5$ on hidden layers 1 & 2



Training Step 1



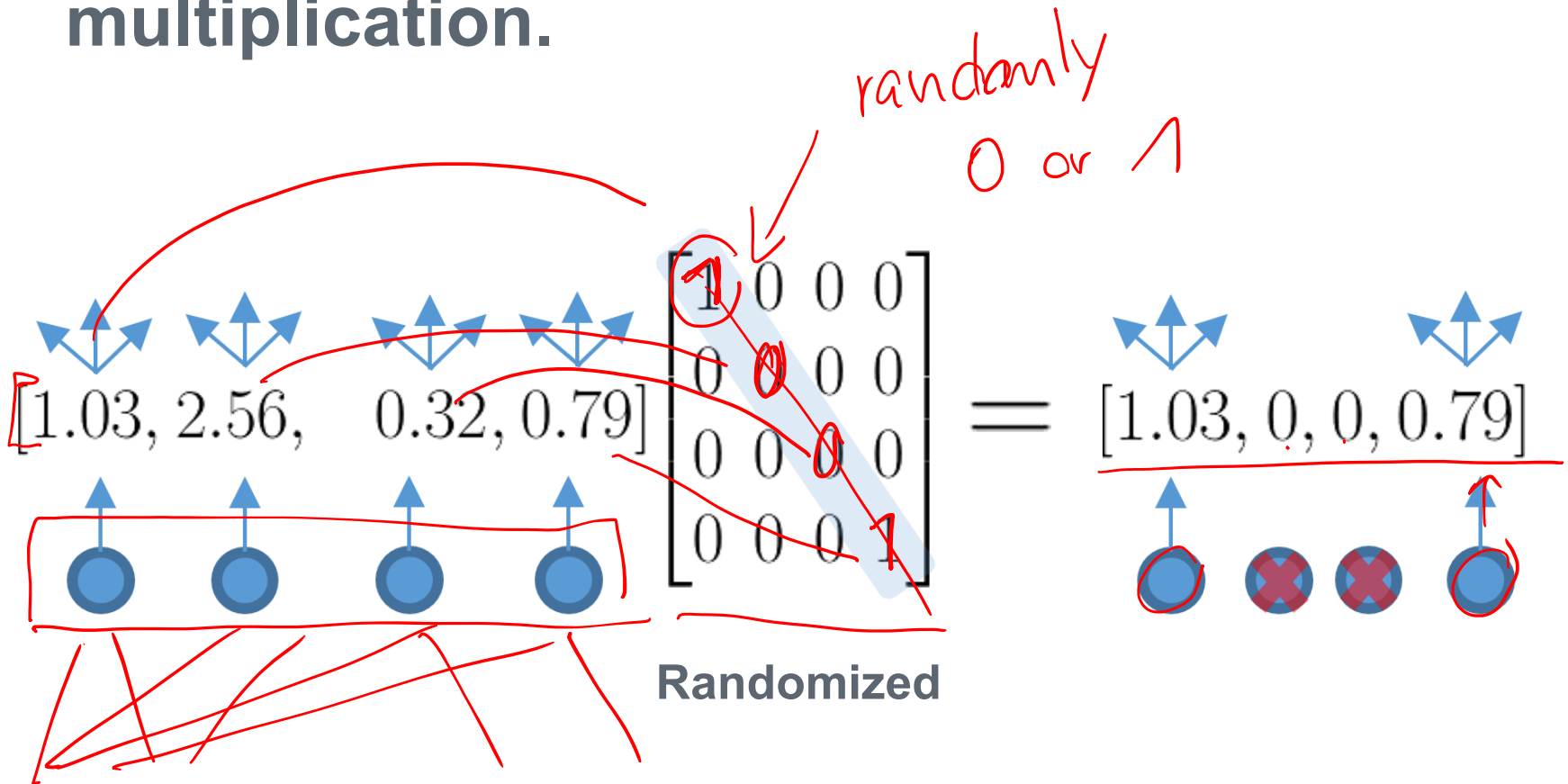
Training Step 2

...

...

Dropping Neurons

This can be done simply through matrix multiplication.



Dropping Neurons

This can be done simply through matrix multiplication.

The diagram illustrates the process of dropping neurons in a neural network layer. It shows a 4x4 matrix multiplication where the first two columns are dropped, resulting in a 1x4 vector where the first two elements are zero.

On the left, a 4x4 matrix is shown with the following values (some in red):

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix is multiplied by a 1x4 vector:

$$[1.03, 2.56, -0.32, 0.79]$$

The result is a 1x4 vector:

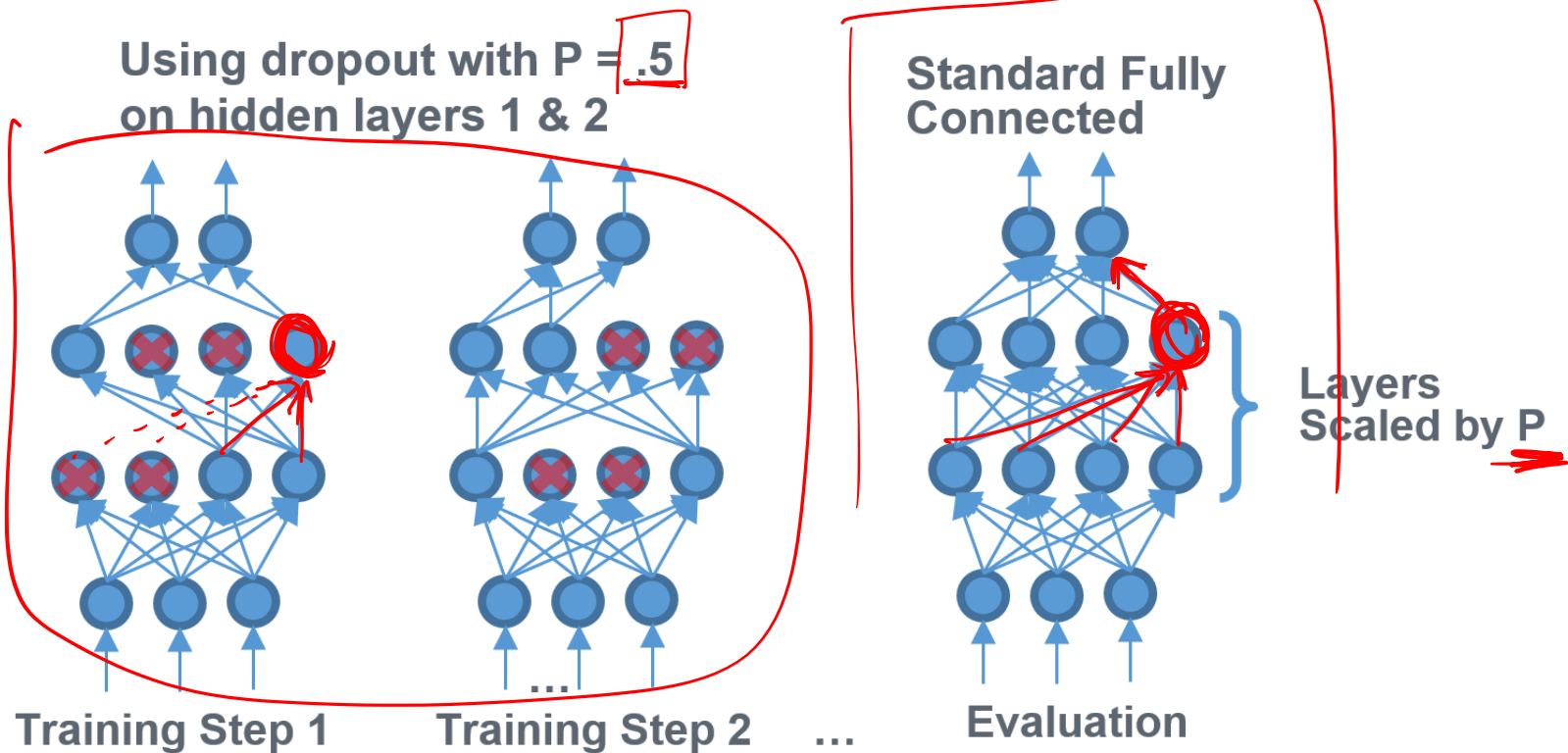
$$[0, 0, -0.32, 0.79]$$

The first two elements of the result vector are zero, indicating that the first two neurons have been dropped. The last two elements, -0.32 and 0.79 , are underlined in red.

Below the matrix, the word "Randomized" is written. Below the result vector, the first two elements are marked with red 'X's and the last two with red checkmarks, indicating which neurons are active.

Dropout Evaluation

- | Dropout is applied during training.
- | During network evaluation, dropout layer neurons are instead scaled by P .



Summary

- | Properly fit model to data
- | Dropout approximates using multiple differently trained networks.
- | Prevents overreliance on specific inputs or combinations
- | Training neurons dropped out with probability P , during evaluation/inference neurons scaled by P .

