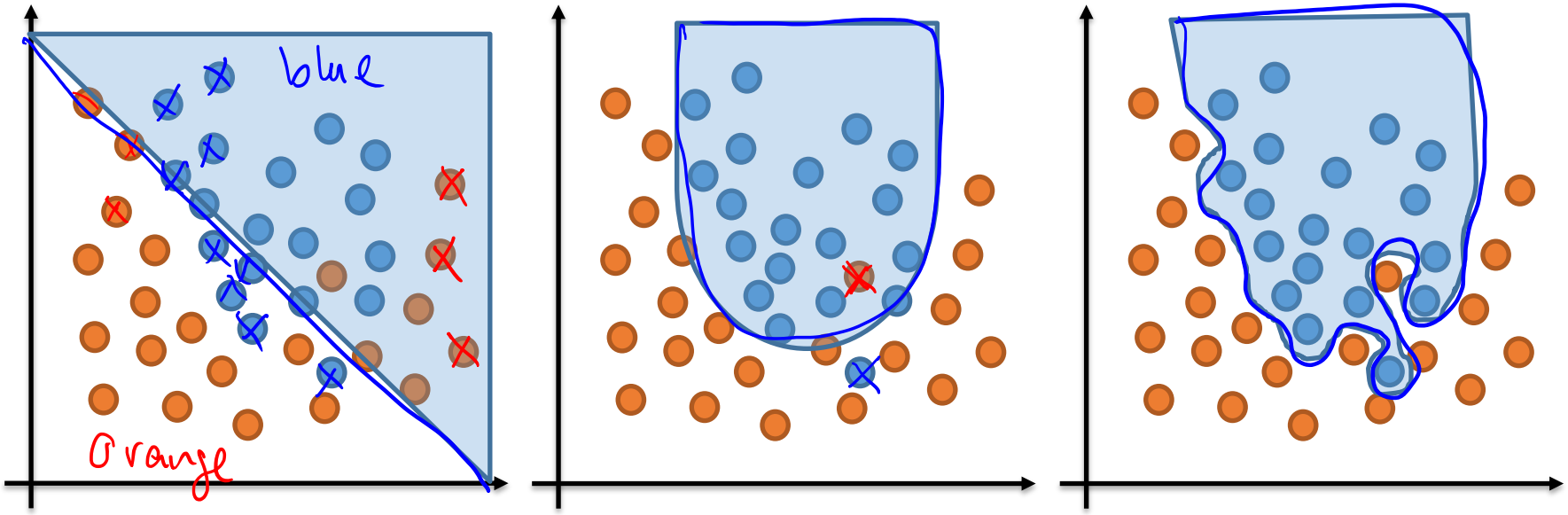
Dropout

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

Ira A. Fulton Schools of Engineering
Arizona State University

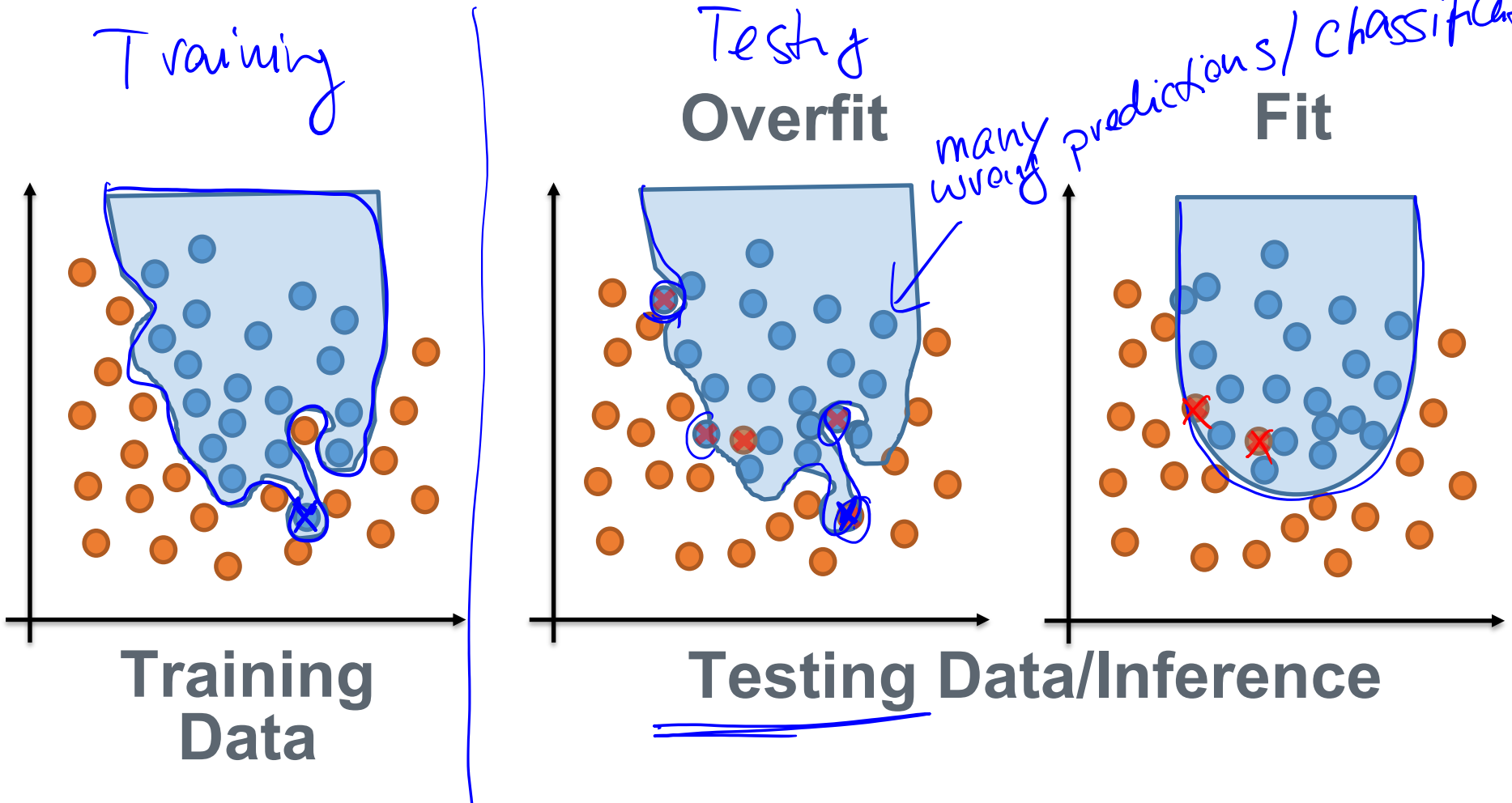
Overfitting

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



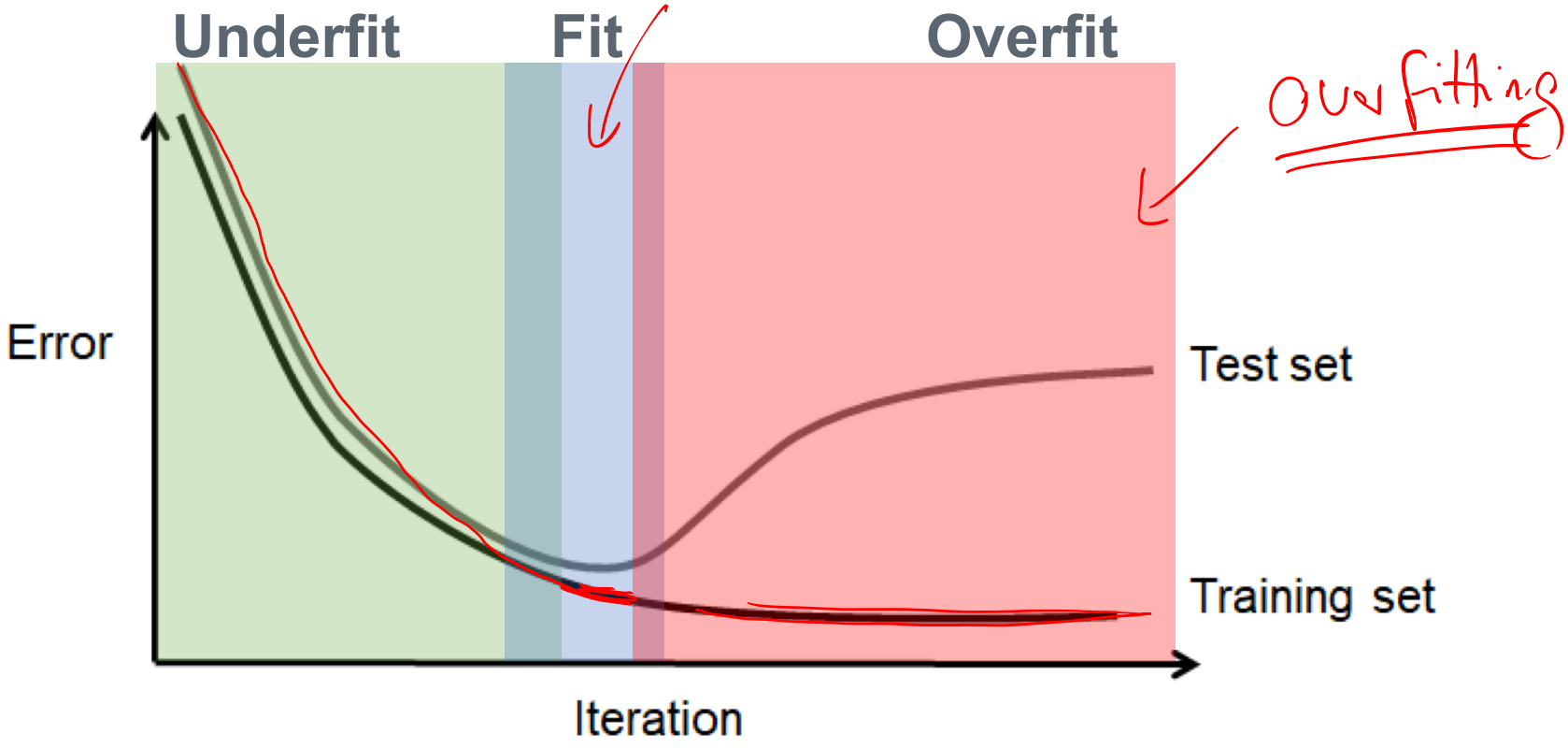
Overfitting

- Worse runtime accuracy on unseen data



Solutions: Early stopping

- Stop training when loss on testing dataset begins increasing.



Solutions: Dropout

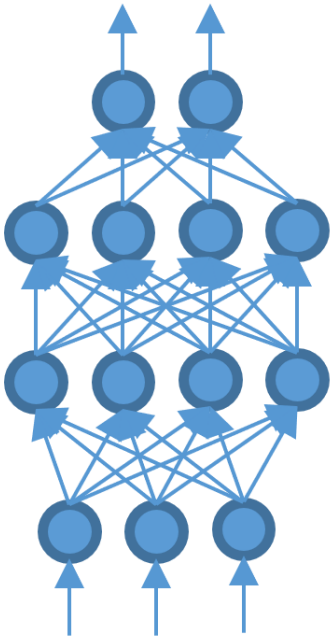
- Introduced by Geoffrey Hinton et al. in 2012
- Mean of many differently trained networks likelier to produce better results.
- Dropout approximates using multiple neural networks without the costly computation or memory.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*

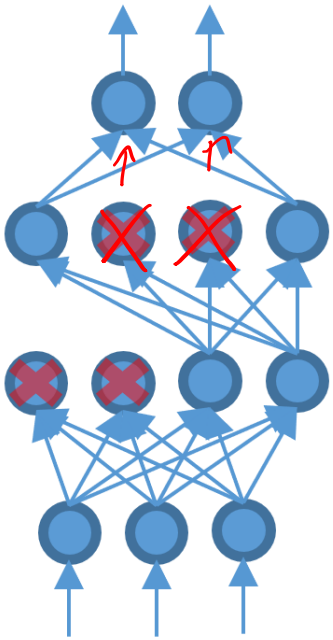
Dropout Training

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

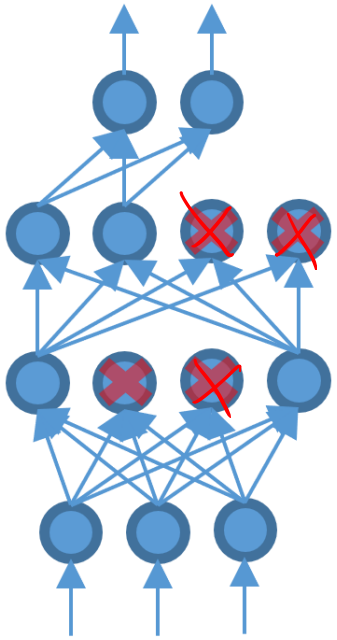
Standard Fully Connected



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



Training Step 1



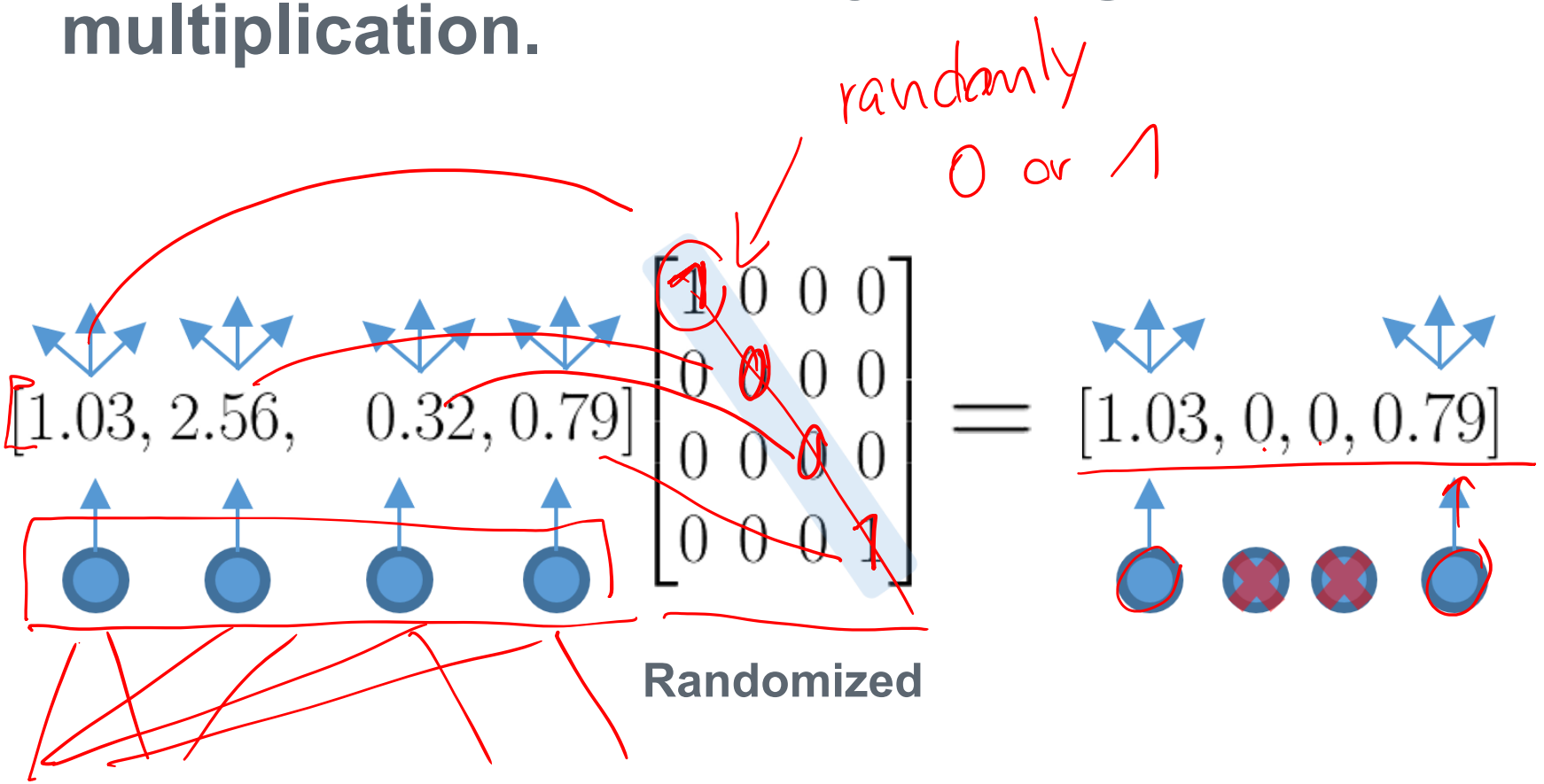
Training Step 2

…

…

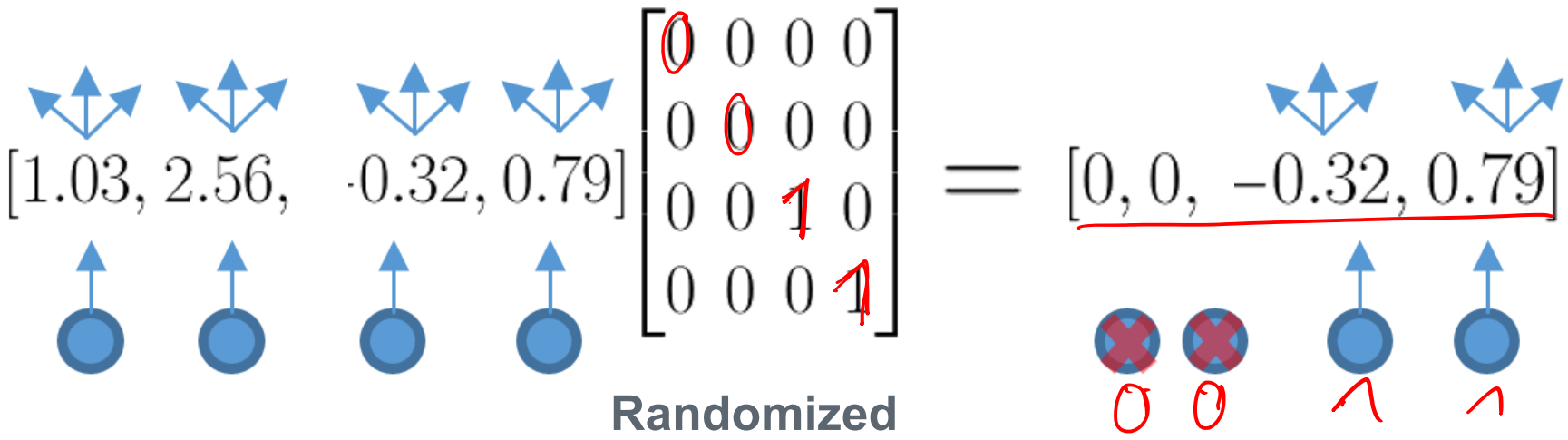
Dropping Neurons

- This can be done simply through matrix multiplication.



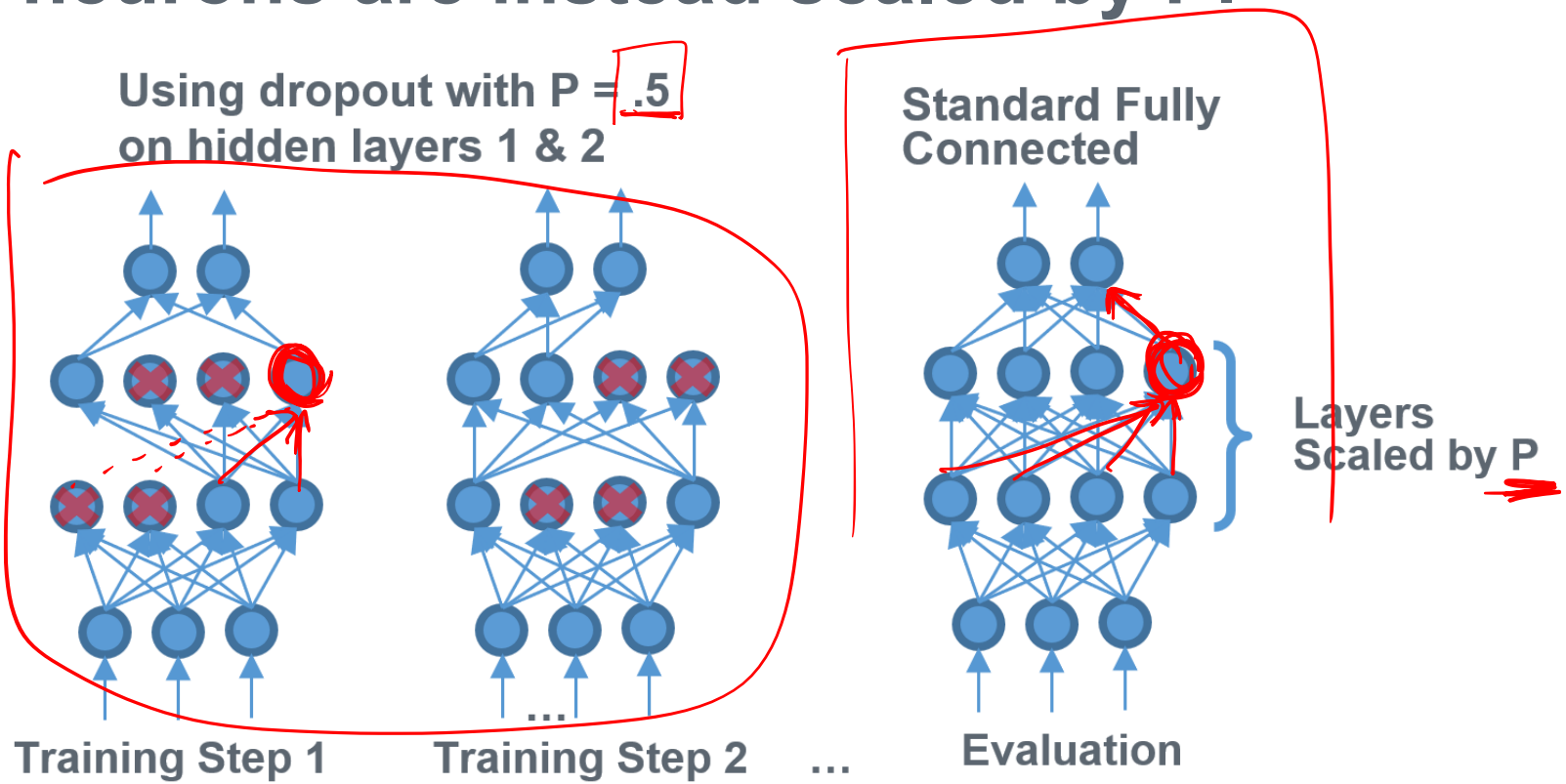
Dropping Neurons

- This can be done simply through matrix multiplication.



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.