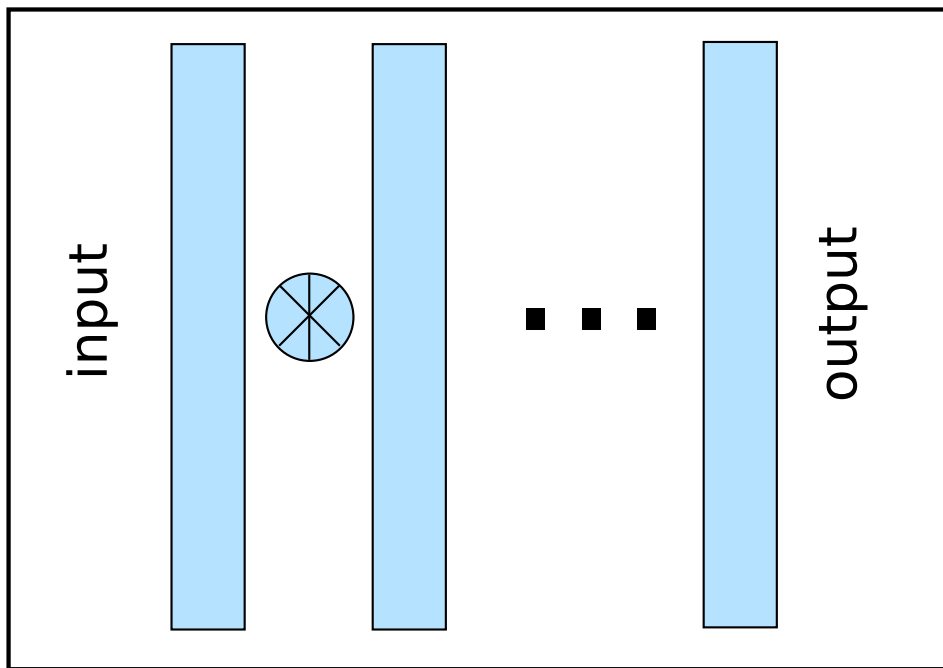Param 1: activation (e.g. tanh, relu)
Param 2: variance of the last layer weight (determines
          the variance of the initial network output)
Param 3 - : layer size (input layer size, hidden layer1
          size, hidden layer2 size ... output layer size)