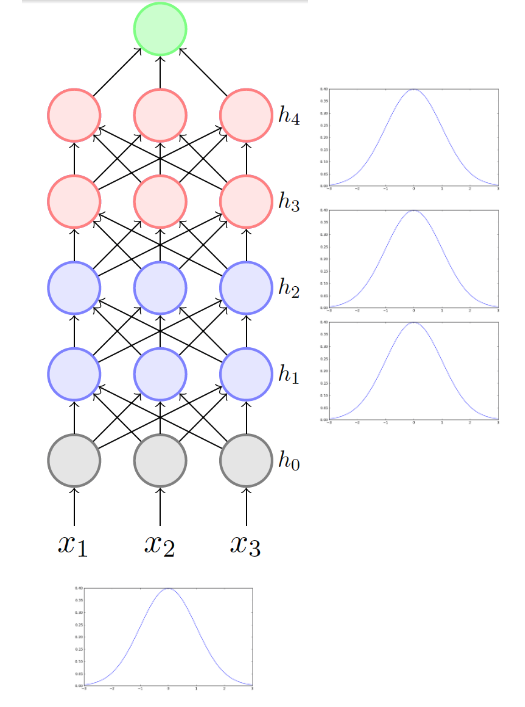
## **Batch Normalization**

We have standardized inputs but what about hidden representations?

1. When we normalize/standardize the inputs, most of the inputs lie around the mean value, in the standard deviation range for all inputs xmn where (m: no. of training samples, n: no. of features)
2. Now, let’s try applying normalizations to any particular hidden representations. Consider one particular Hidden representation H
3. Now, standardization is done on the H matrix so that all of the columns are in the same range.
4. In a sense, the H matrix acts as an input to the next layer.
5. Thus, to learn the weights effectively, Standardization is done on all the columns to bring them to a comparable range.
6. Just as we standardize the inputs, we standardize the activations at all layers
7. Now, why is it called batch normalization?
8. Let’s take a look at the formulae for Batch normalizing H
   1. for each feature of each training sample.
   2. Where mean:
   3. And Variance/Standard deviation:
9. Now, for the input normalization, we consider the entire training size (m) when computing mean and std. However, here we only consider a smaller subset of samples (k). **Mean (μ) and Std (σ) are calculated using a batch of k samples, hence it is called batch normalization.** K = 32 is commonly used.