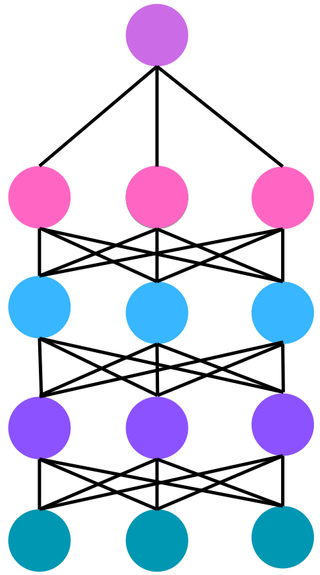
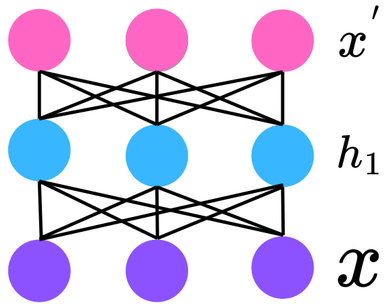
Un SupervisedPretraining

- we use auto-encoder for each input and consecutive hidden layers to learn feature
- we repeat this unsupervised learning to learn all the hidden layers upto last hidden layer
- After this layerwise pretraining we can add the output layer and the training the whole network using the task specific objective

$$min\frac{1}{m}\sum(y_i - f(x_i))^2$$





- Is it a better optimizer?
- Is it a regularizor?
- If DNN is of large capacity we don't need to pretrain
- But if we have deep but not large network then we need pretraining.
- Pretraining conatrints the weights to lie in only certain regions of the parameter space
- It conatsraints the weights to lie in regions where the characteristics of the data are captured well