

[3] Contrastive Divergence ^{like algo} can be used for efficient training.

Persuade code is as follows:

init $y^{(0)}, z^{(0)}$, and use training set

$$y^{(1)} \sim P(y | x^{(0)}, z^{(0)})$$

$$z^{(1)} \sim P(z | x^{(0)}, y^{(1)})$$

and,