$Decoder: \hat{x} = sigm(c + W^* h(x))\ (sigm\ for\ binary\ inputs)$

$Encoder: h(x) = sigm(b + Wx)$

$with\ SVD: A = U\, \Sigma\, \mathrm{V}^{\mathrm{T}}$

$we\ keep\ only\ the\ k\ largest\ singular\ values\ SVD, matrix\ B\ of\ rank\ k\ is\ closest\ to\ A$

$$B^* = \underset{rank\ (B)=k}{\operatorname{argmin}} \|A - B\|_F$$
$$B^* = U_{.,\leq k}\, \Sigma_{,\leq k,\leq k}\, \mathrm{V}^{\mathrm{T}}_{.,\leq k}$$

$$\min_{\theta} \sum_t \frac{1}{2} \sum_i \left(x_i^{(t)} - \hat{x}_i^{(t)}\right)^2 \geq \min_{W^*, h(x)} \frac{1}{2} \|X - W^* h(X)\|_F^2$$

$$\underset{W^*, h(x)}{\operatorname{argmin}} \frac{1}{2} \|X - W^* h(X)\|_F^2 = (W^* \leftarrow U_{.,\leq k} \Sigma_{\leq k,\leq k}\ , h(X) \leftarrow V^T_{.,\leq k})$$

$$h(X) = V^T_{.,\leq k}$$
$$h(X) = V^T_{.,\leq k}\ (X^T\ X)^{-1} (X^T\ X)$$
$$h(X) = V^T_{.,\leq k}\ \left(V\, \Sigma^{\mathrm{T}} U^T\ U\, \Sigma\, \mathrm{V}^{\mathrm{T}}\right)^{-1} \left(V\, \Sigma^{\mathrm{T}} U^T\ X\right)$$
$$h(X) = V^T_{.,\leq k}\ \left(\mathrm{V}\, \Sigma^{\mathrm{T}}\ \ \Sigma\, \mathrm{V}^{\mathrm{T}}\right)^{-1} V\, \Sigma^{\mathrm{T}} U^T\ X$$
$$h(X) = V^T_{.,\leq k}\ V \left(\Sigma^{\mathrm{T}}\ \ \Sigma\right)^{-1} \mathrm{V}^{\mathrm{T}}\ \ V\, \Sigma^{\mathrm{T}} U^T\ X$$
$$h(X) = V^T_{.,\leq k}\ V \left(\Sigma^{\mathrm{T}}\ \ \Sigma\right)^{-1} \Sigma^{\mathrm{T}} U^T\ X$$
$$h(X) = I^T_{\leq k}\ \left(\Sigma^{\mathrm{T}}\ \ \Sigma\ \right)^{-1} \Sigma^{\mathrm{T}} U^T\ X$$
$$h(X) = I^T_{\leq k}\ \Sigma^{-1} \left(\Sigma^{\mathrm{T}}\right)^{-1} \Sigma^{\mathrm{T}} U^T\ X$$
$$h(X) = I^T_{\leq k}\ \Sigma^{-1} U^T\ X$$
$$h(X) = \Sigma^{-1}_{\leq k,\leq k} (U_{.,\leq k})^T\ X$$