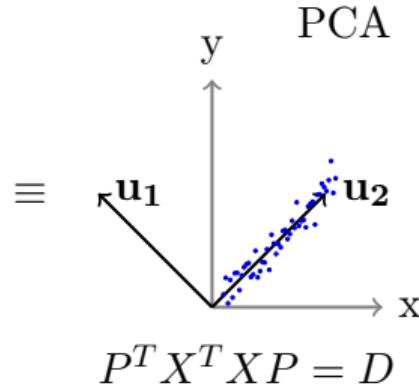
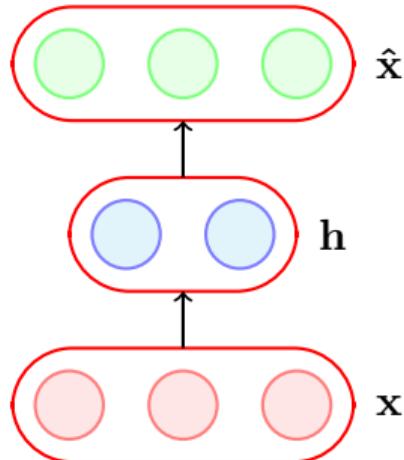
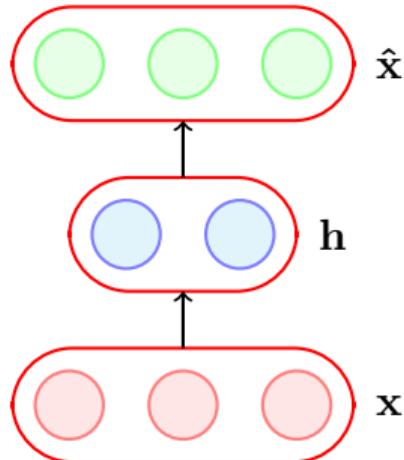


## Module 7.7 : Summary



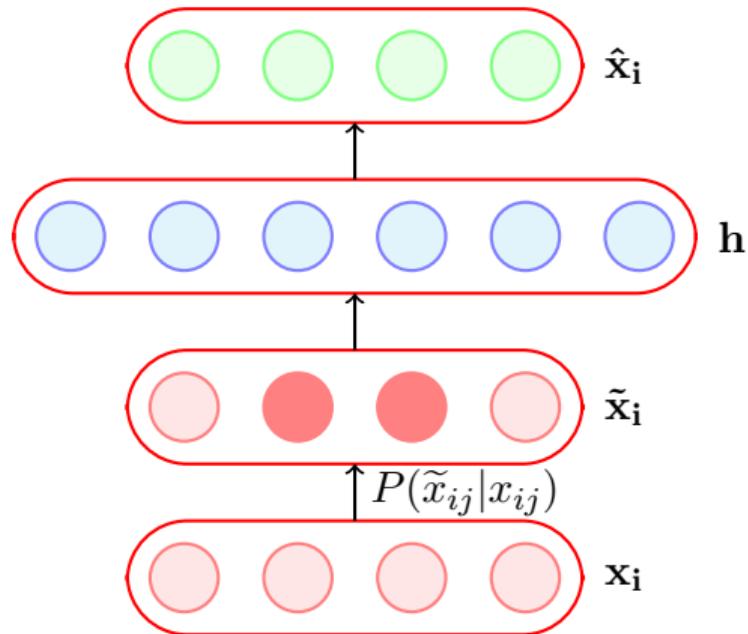


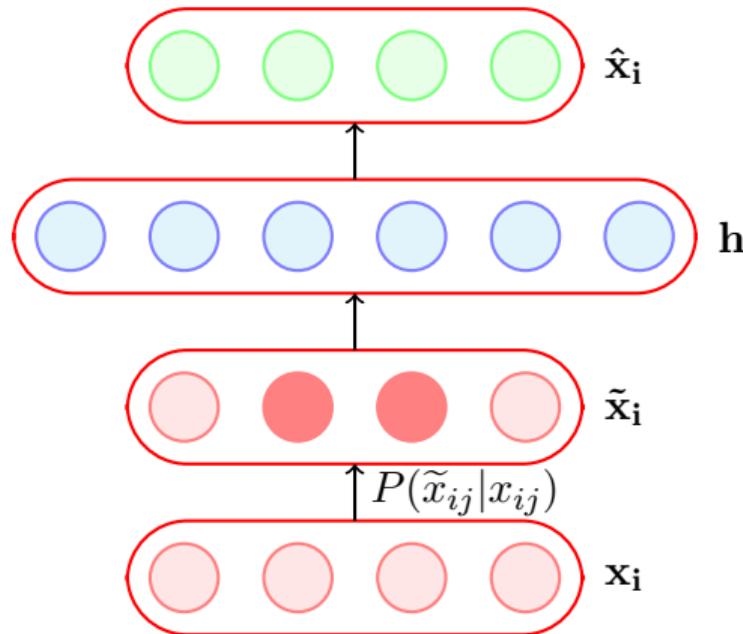
$$\hat{\mathbf{x}} \quad \text{PCA} \\ \mathbf{x} \equiv \begin{matrix} \mathbf{u}_1 & \mathbf{u}_2 \\ \mathbf{y} & \mathbf{x} \end{matrix}$$

The diagram shows a PCA decomposition. A vector  $\mathbf{x}$  is represented as a linear combination of orthogonal basis vectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$ , with a scalar coefficient  $y$ . This is equivalent to the neural network structure shown above.

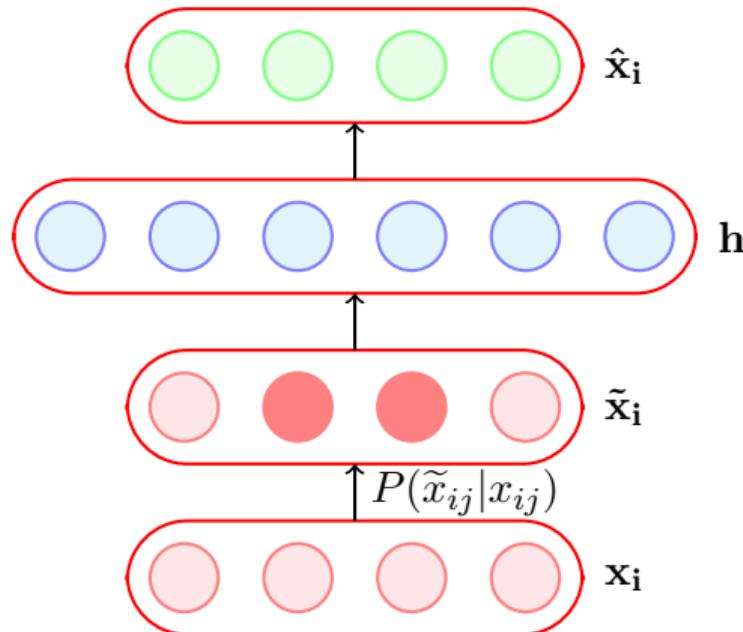
$$P^T X^T X P = D$$

$$\min_{\theta} \|X - \underbrace{HW^*}_{{U\Sigma V^T} \atop (\text{SVD})}\|_F^2$$





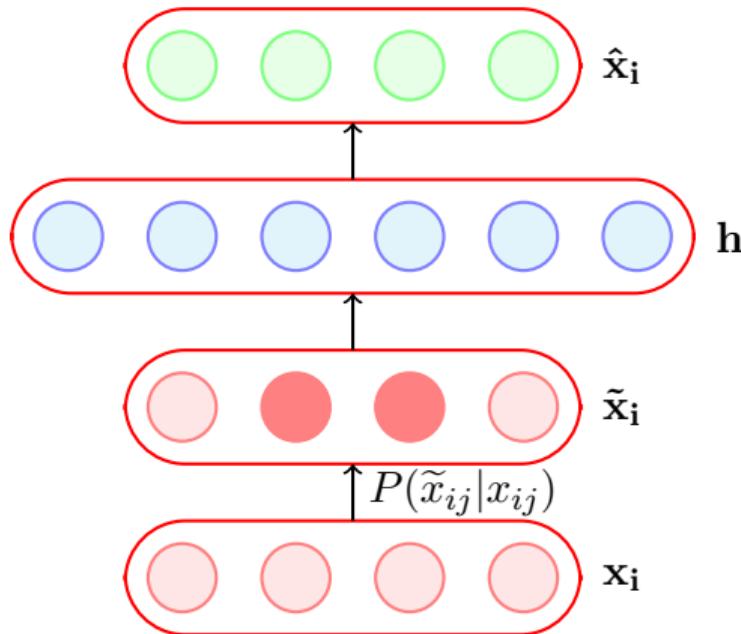
Regularization



Regularization

$$\Omega(\theta) = \lambda \|\theta\|^2$$

Weight decaying



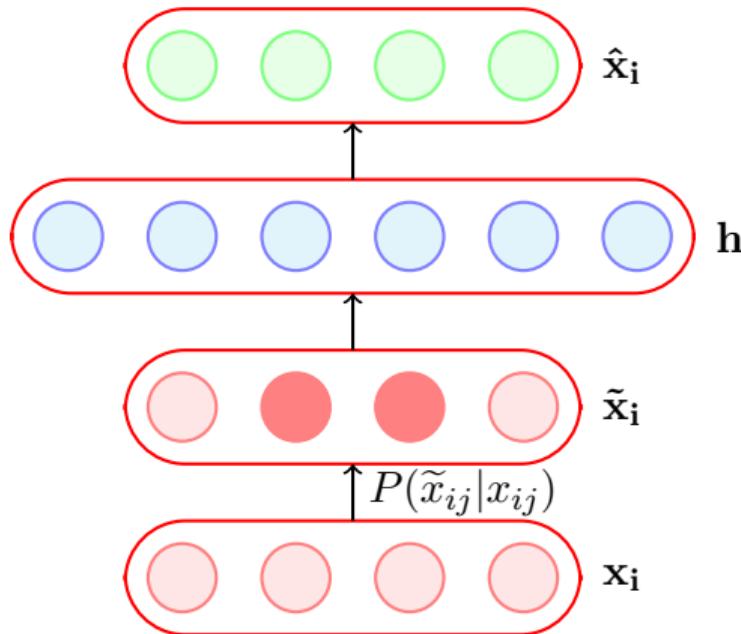
## Regularization

$$\Omega(\theta) = \lambda \|\theta\|^2$$

Weight decaying

$$\Omega(\theta) = \sum_{l=1}^k \rho \log \frac{\rho}{\hat{\rho}_l} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_l}$$

Sparse



## Regularization

$$\Omega(\theta) = \lambda \|\theta\|^2$$

Weight decaying

$$\Omega(\theta) = \sum_{l=1}^k \rho \log \frac{\rho}{\hat{\rho}_l} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_l}$$

Sparse

$$\Omega(\theta) = \sum_{j=1}^n \sum_{l=1}^k \left( \frac{\partial h_l}{\partial x_j} \right)^2$$

Contractive