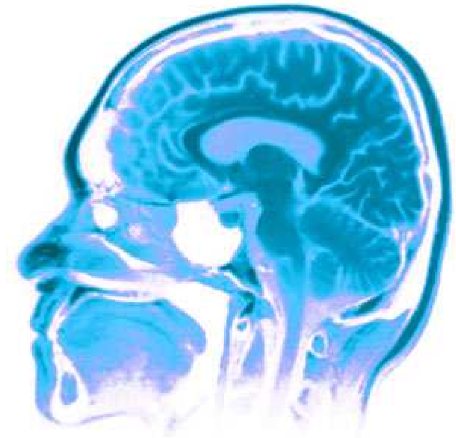
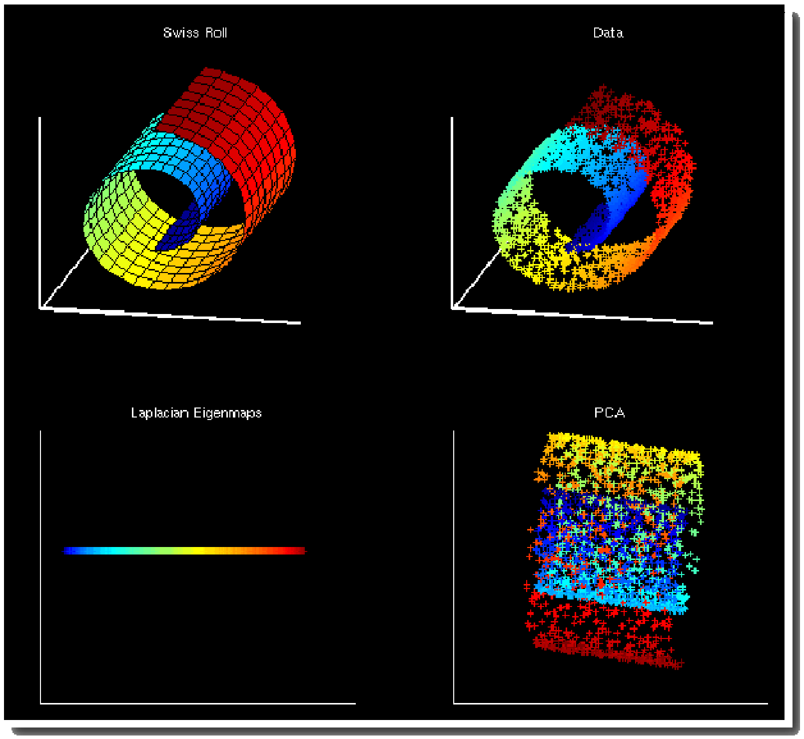
CPSC540

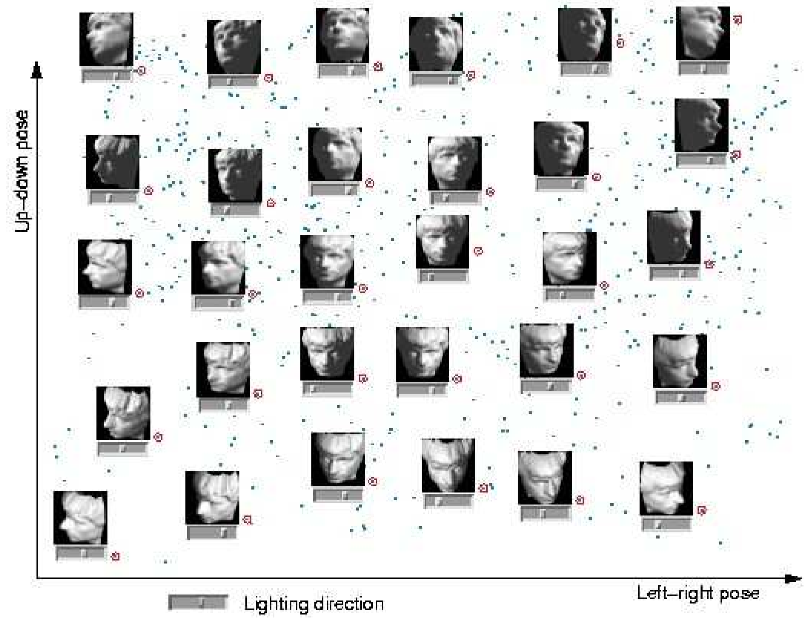


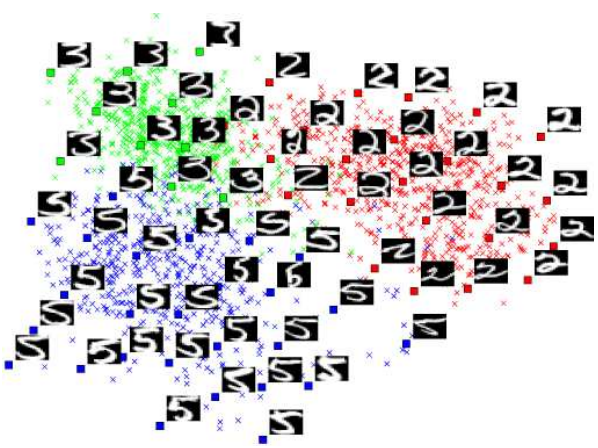
Principal Component Analysis (PCA)

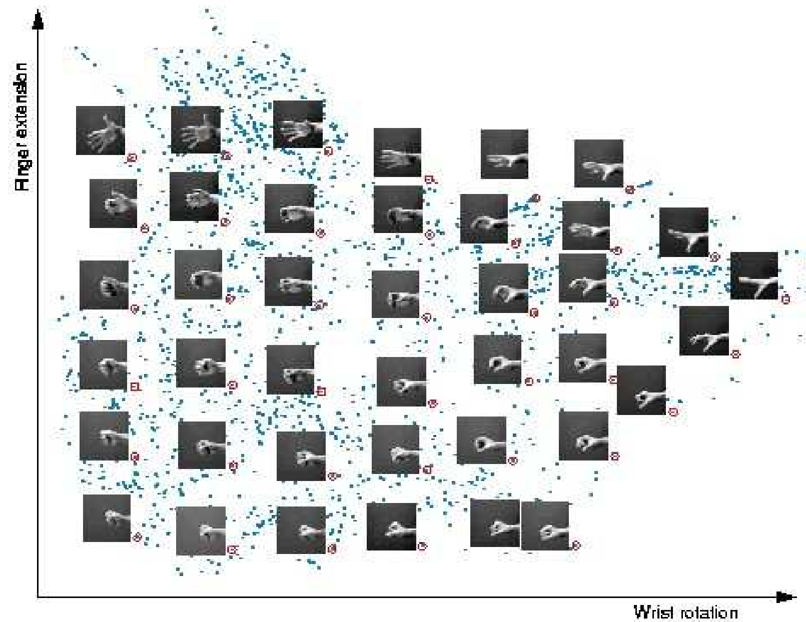
Nando de Freitas
September, 2011
University of British Columbia

Principal component analysis (PCA)









PCA derivation: 2D to 1D

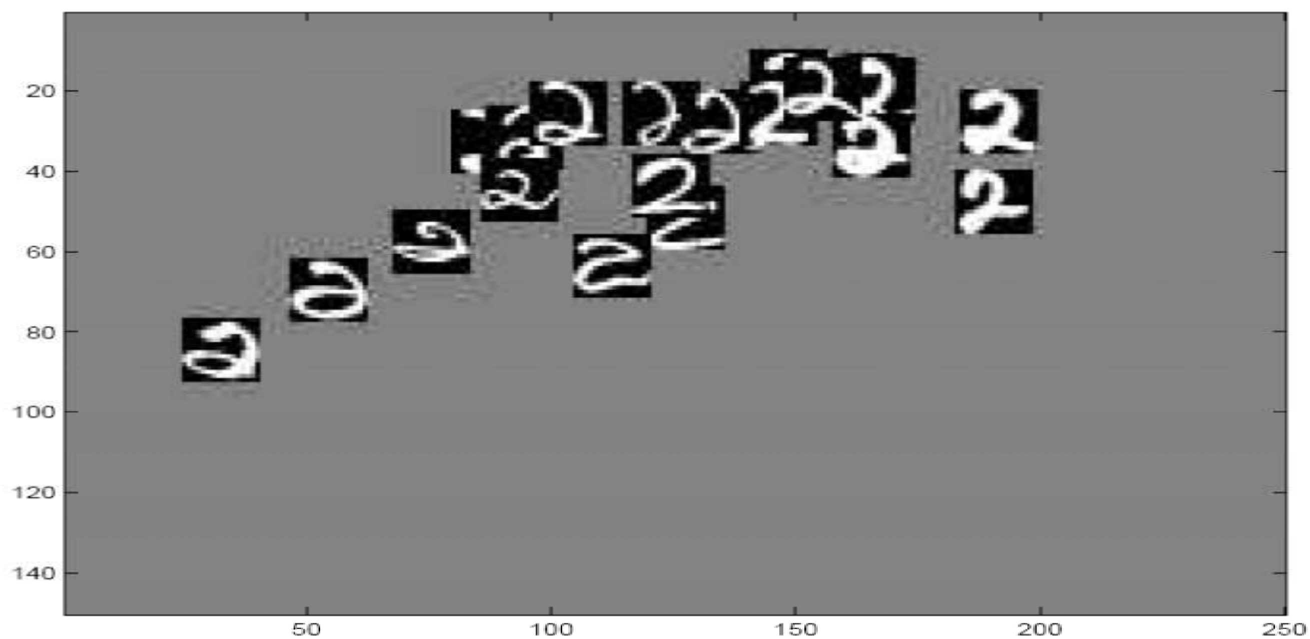
$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{i1} & x_{i2} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ \vdots & \vdots \\ u_{i1} & u_{i2} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix}$$

$$\mathbf{x}_i = u_{i1}\sigma_1\mathbf{v}_1^T + u_{i2}\sigma_2\mathbf{v}_2^T$$

$$\widehat{\mathbf{x}}_i = u_{i1}\sigma_1\mathbf{v}_1^T$$

PCA visualization

For example, we can take several 16×16 images of the digit 2 and project them to 2D. The images can be written as vectors with 256 entries. We then from the matrix $\mathbf{A} \in \mathbb{R}^{n \times 256}$, carry out the SVD and truncate it to $k = 2$. Then the components $\mathbf{U}_k \mathbf{\Sigma}_k$ are 2 vectors with n data entries. We can plot these 2D points on the screen to visualize the data.



PCA in python

```
>>> U, S, V = svd(A)
>>> k = 2
>>> Z = dot(U[:,:k], eye(k) * S[:k])
>>> figure(1)
>>> plot(Z[:,0], Z[:,1], 'ro')
>>> grid()
```

PCA as orthogonal reconstruction

data matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$

basis vectors $\mathbf{w}_j \in \mathbb{R}^{1 \times d}$

scores $\mathbf{z}_i \in \mathbb{R}^{1 \times l}$

we minimize the average reconstruction error:

$$J(\mathbf{W}, \mathbf{Z}) = \frac{1}{n} \sum_{i=1}^{n} ||\mathbf{x}_i - \hat{\mathbf{x}}_i||^2 = \frac{1}{n} \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{z}_i \mathbf{W}||^2 = ||\mathbf{X} - \mathbf{Z}\mathbf{W}||_F^2$$

subject to the constraint that $\mathbf{W} \in \mathbb{R}^{l \times d}$ has orthonormal bases $\mathbf{w}_j$.

Standardize the data first!

$$x_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_j}$$

$$\overline{x}_j = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_{ij}$$

$$\sigma_j^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$$

The first component

$$J(\mathbf{w}_1, \mathbf{z}_1) = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{x}_i - z_{i1}\mathbf{w}_1||^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - z_{i1}\mathbf{w}_1)(\mathbf{x}_i - z_{i1}\mathbf{w}_1)^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\mathbf{x}_i\mathbf{x}_i^T - 2z_{i1}\mathbf{w}_1\mathbf{x}_i^T + z_{i1}^2\mathbf{w}_1\mathbf{w}_1^T]$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\mathbf{x}_i\mathbf{x}_i^T - 2z_{i1}\mathbf{w}_1\mathbf{x}_i^T + z_{i1}^2]$$

$$\frac{\partial}{\partial z_{i1}} J(\mathbf{w}_1, \mathbf{z}_1) = \frac{1}{n}[-2\mathbf{w}_1\mathbf{x}_i^T + 2z_{i1}] = 0 \Rightarrow z_{i1} = \mathbf{w}_1\mathbf{x}_i^T$$

$$J(\mathbf{w}_1) = \frac{1}{n}\sum_{i=1}^{n}[\mathbf{x}_i\mathbf{x}_i^T - z_{i1}^2] = \text{const} - \frac{1}{n}\sum_{i=1}^{n} z_{i1}^2$$

min reconstruction = max variance

$$J(\mathbf{w}_1) = \frac{1}{n}\sum_{i=1}^{n}[\mathbf{x}_i\mathbf{x}_i^T - z_{i1}^2] = \text{const} - \frac{1}{n}\sum_{i=1}^{n} z_{i1}^2$$

$\tilde{\mathbf{z}}_j \in \mathbb{R}^n$ to denote the j'th column of $\mathbf{Z}$

$$\text{var}\left[\tilde{\mathbf{z}}_1\right] = \mathbb{E}\left[\tilde{\mathbf{z}}_1^2\right] - \left(\mathbb{E}\left[\tilde{\mathbf{z}}_1\right]\right)^2 = \frac{1}{n}\sum_{i=1}^{n} z_{i1}^2 - 0$$

$$\mathbb{E}\left[z_{i1}\right] = \mathbb{E}\left[\mathbf{w}_1\mathbf{x}_i^T\right] = \mathbf{w}_1\mathbb{E}\left[\mathbf{x}_i\right]^T = 0$$

$$\arg\min_{\mathbf{w}_1} J(\mathbf{w}_1) = \arg\max_{\mathbf{w}_1} \text{var}\left[\tilde{\mathbf{z}}_1\right]$$

Variance of projections

$$\frac{1}{n}\sum_{i=1}^{n} z_{i1}^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_1 \mathbf{x}_i^T \mathbf{x}_i \mathbf{w}_1^T = \mathbf{w}_1 \hat{\mathbf{\Sigma}} \mathbf{w}_1^T$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i$$

impose the constraint $||\mathbf{w}_1|| = 1$

maximize the following objective:

$$\tilde{J}(\mathbf{w}_1) = \mathbf{w}_1^T \hat{\mathbf{\Sigma}} \mathbf{w}_1 + \lambda_1 (1 - \mathbf{w}_1 \mathbf{w}_1^T)$$

Lagrange multiplier λ_1

Spectral analysis

$$\frac{\partial}{\partial \mathbf{w}_1}\tilde{J}(\mathbf{w}_1) = 2\hat{\mathbf{\Sigma}}\mathbf{w}_1 - 2\lambda_1\mathbf{w}_1 = 0$$

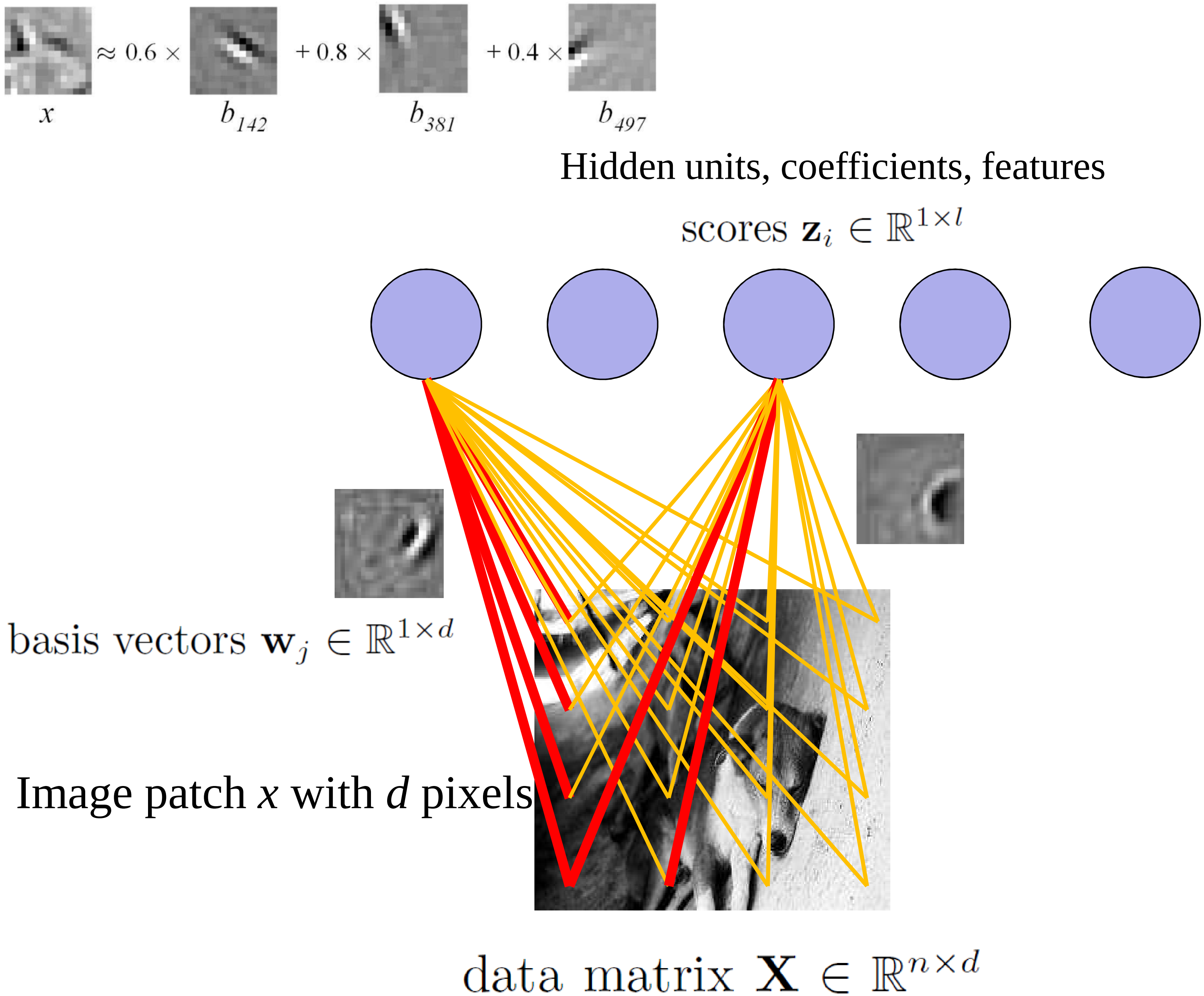
$$\hat{\mathbf{\Sigma}}\mathbf{w}_1 = \lambda_1\mathbf{w}_1.$$

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

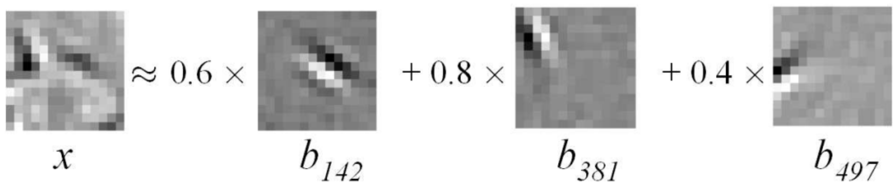
$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{X}^T\mathbf{X} = \frac{1}{n}\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T \qquad \mathbf{W} = \mathbf{V}$$

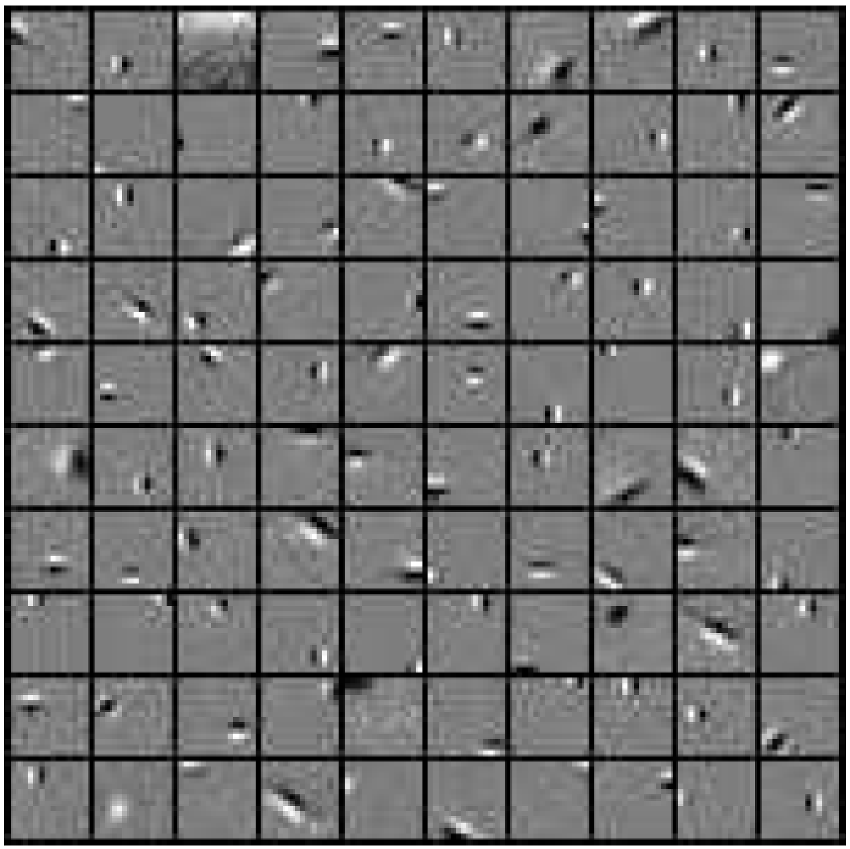
$$z_{ij} = \mathbf{w}_j\mathbf{x}_i^T$$

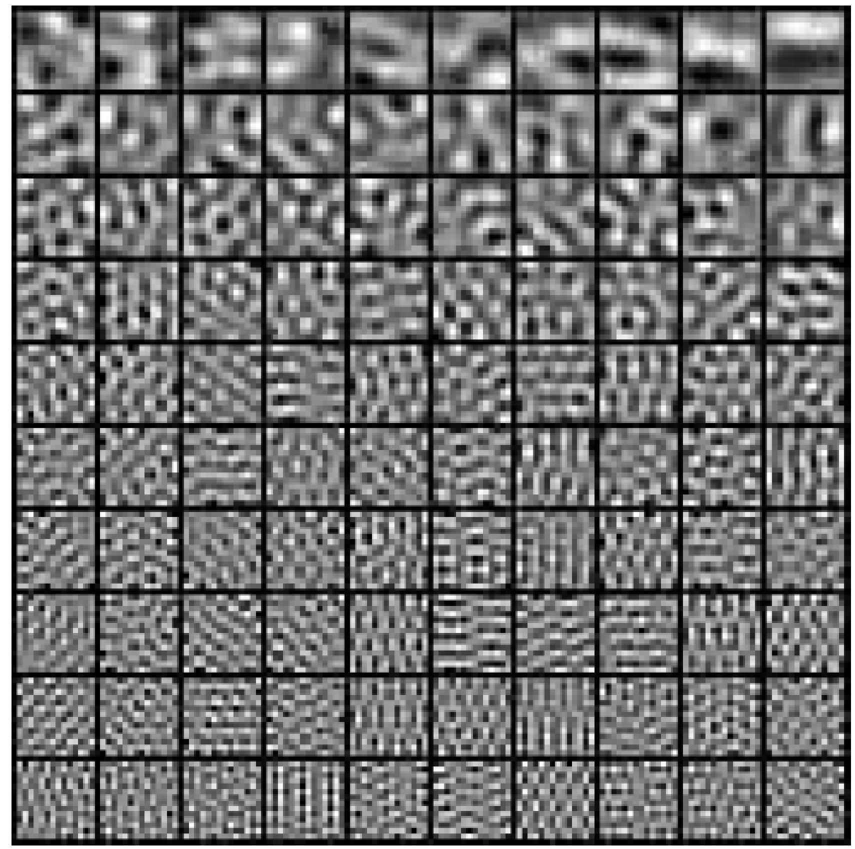
$$\mathbf{Z} = \mathbf{X}\mathbf{W}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$$


$x \approx 0.6 \times b_{142} + 0.8 \times b_{381} + 0.4 \times b_{497}$
Hidden units, coefficients, features
scores $\mathbf{z}_i \in \mathbb{R}^{1\times l}$
basis vectors $\mathbf{w}_j \in \mathbb{R}^{1\times d}$
Image patch x with d pixels
data matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$

PCA bases for image patches





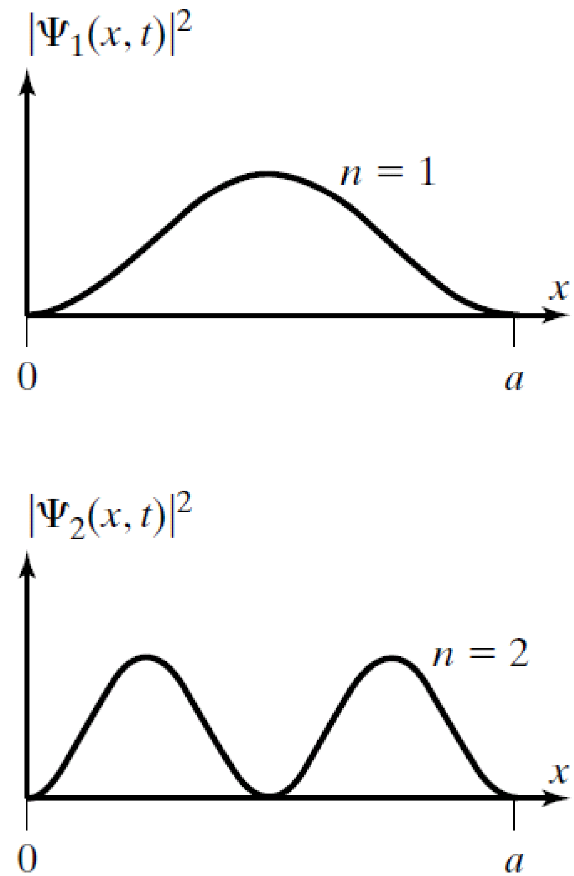


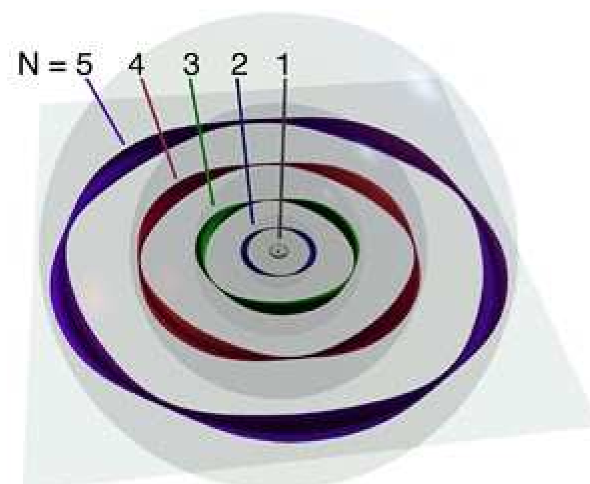
$$\mathbf{B}^*, \mathbf{C}^* = \arg\min_{\mathbf{B},\mathbf{C}} ||\mathbf{X} - \mathbf{BC}||_2^2 + \lambda \, \|\mathbf{C}\|_1$$

$$s.t. \quad ||\mathbf{b}_j||_2^2 = 1, \ \forall j.$$

Schrodinger's equation

$$E\psi(x) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + U(x)\right)\psi(x)$$





Bohr-de Broglie electron matterwave orbits shells 1-5