

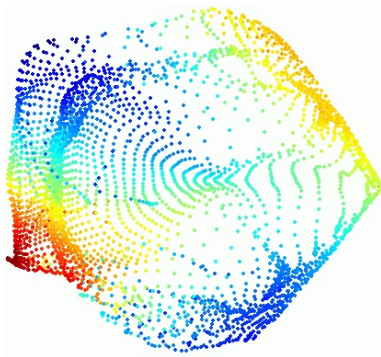
Data Representation as Low Rank Matrix Factorization

Ziv Epstein
`ziv.epstein@pomona.edu`

Pomona College

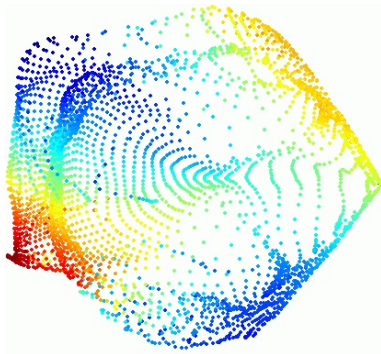
February 13, 2017

In many contexts in data science and linear algebra, we have lots of points in a high-dimensional space.



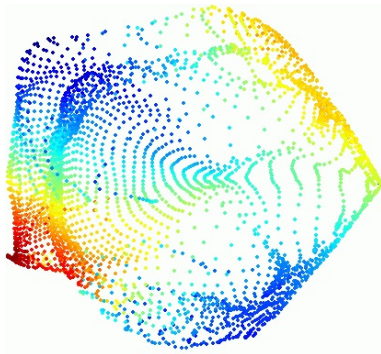
Denote these $x_i \in \mathbb{R}^m$ for $i = 1 \dots n$.

How do we cluster x_i ?



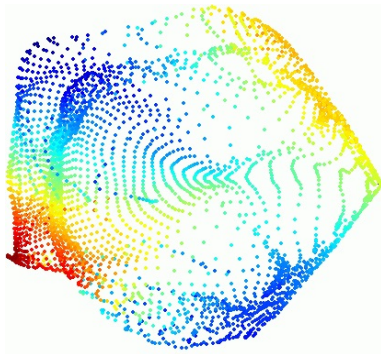
Denote these $x_i \in \mathbb{R}^m$ for $i = 1 \dots n$.

How do we cluster x_i ? How do we perform dimensionality reduction?



Denote these $x_i \in \mathbb{R}^m$ for $i = 1 \dots n$.

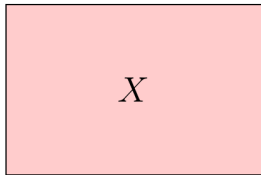
How do we cluster x_i ? How do we perform dimensionality reduction? How do we visualize them?



Denote these $x_i \in \mathbb{R}^m$ for $i = 1 \dots n$.

Representation and Factorization

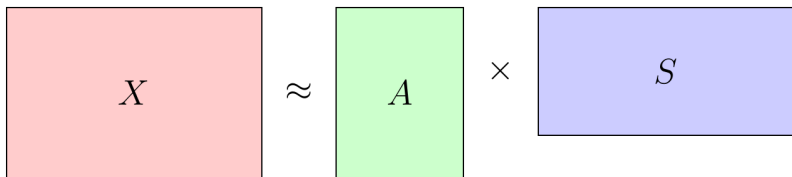
We can *approximate* $X \in \mathbb{R}^{n \times m}$.



with $\text{rank}(X) = \min(m, n)$.

Representation and Factorization

We can *approximate* $X \in \mathbb{R}^{n \times m}$.



with $\text{rank}(AS) = k \ll \min(m, n)$.

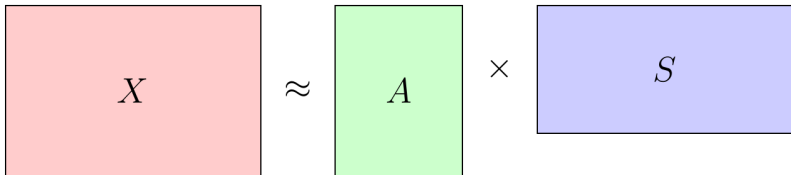
Our original high-dimensional points are linear combinations of “basis elements’.

$$X_i = \sum_{j=1}^k A_{i,j} S_j$$

Low Dimensional Interpretation

Our original high-dimensional points are linear combinations of “basis elements”.

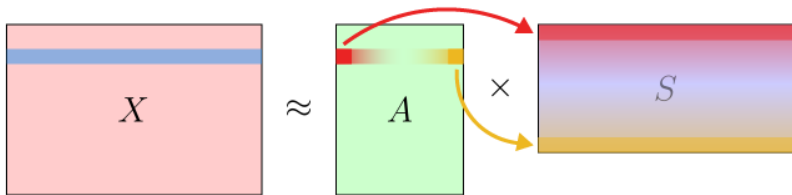
$$X_i = \sum_{j=1}^k A_{i,j} S_j$$



Low Dimensional Interpretation

Our original high-dimensional points are linear combinations of “basis elements”.

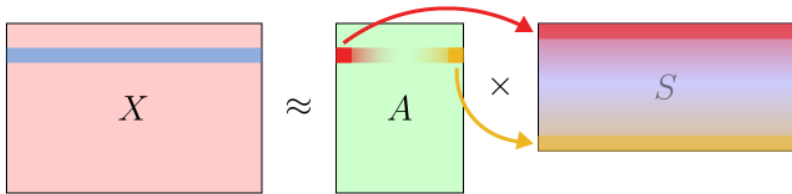
$$X_i = \sum_{j=1}^k A_{i,j} S_j$$



Low Dimensional Interpretation

Our original high-dimensional points are linear combinations of “basis elements”.

$$X_i = \sum_{j=1}^k A_{i,j} S_j$$



Find factorization by solving optimization problem

$$\min_{A,S} ||X - AS||$$

Example (Lee and Seung 1999)

Example (Lee and Seung 1999)

A 19x19 pixel greyscale image of a face is a data point ($x_i \in \mathbb{R}^{361}$ with values between 0 and 256)

Example (Lee and Seung 1999)

A 19x19 pixel greyscale image of a face is a data point ($x_i \in \mathbb{R}^{361}$ with values between 0 and 256)



Example (Lee and Seung 1999)

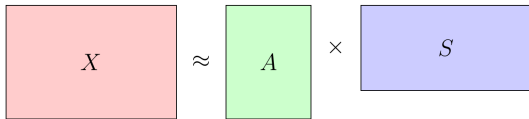
A 19x19 pixel greyscale image of a face is a data point ($x_i \in \mathbb{R}^{361}$ with values between 0 and 256)



Take a database of 2,429 faces to get $X \in \mathbb{R}^{2429 \times 361}$

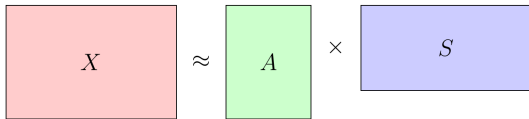
Example (Lee and Seung 1999)

Before: Find any A and S such that $\|X - AS\|$ is minimized.



Example (Lee and Seung 1999)

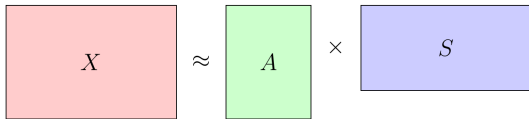
Before: Find any A and S such that $\|X - AS\|$ is minimized.



Now: We add some constraints

Example (Lee and Seung 1999)

Before: Find any A and S such that $\|X - AS\|$ is minimized.

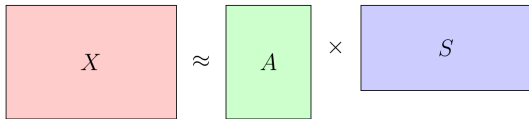

$$X \approx A \times S$$

Now: We add some constraints

- ① Columns of A to be orthonormal; Rows of S to be orthogonal (PCA)

Example (Lee and Seung 1999)

Before: Find any A and S such that $\|X - AS\|$ is minimized.

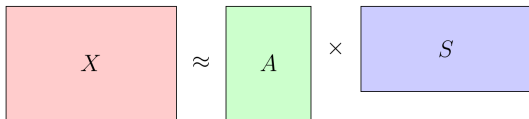

$$X \approx A \times S$$

Now: We add some constraints

- ① Columns of A to be orthonormal; Rows of S to be orthogonal (PCA)
- ② Row of A sums to 1, which one entry equal to one and the rest equal to zero (VQ)

Example (Lee and Seung 1999)

Before: Find any A and S such that $\|X - AS\|$ is minimized.

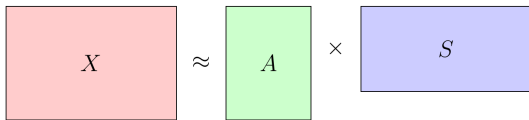

$$X \approx A \times S$$

Now: We add some constraints

- ① Columns of A to be orthonormal; Rows of S to be orthogonal (PCA)
- ② Row of A sums to 1, which one entry equal to one and the rest equal to zero (VQ)
- ③ All entries of A and S are non-negative (NNMF)

Example (Lee and Seung 1999)

Before: Find any A and S such that $\|X - AS\|$ is minimized.


$$X \approx A \times S$$

Now: We add some constraints

- ① Columns of A to be orthonormal; Rows of S to be orthogonal (PCA)
- ② Row of A sums to 1, which one entry equal to one and the rest equal to zero (VQ)
- ③ All entries of A and S are non-negative (NNMF)

For each, find A and S subject to constraints that minimize

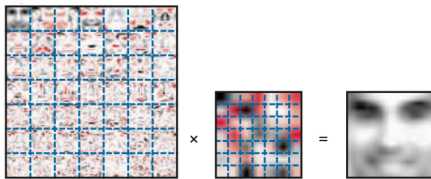
$$\|X - AS\|$$

Example (Lee and Seung 1999)

1: PCA

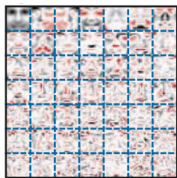
Example (Lee and Seung 1999)

1: PCA



Example (Lee and Seung 1999)

1: PCA



\times



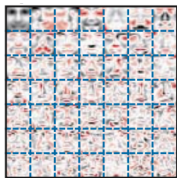
$=$



“eigenfaces”

Example (Lee and Seung 1999)

1: PCA



x



=

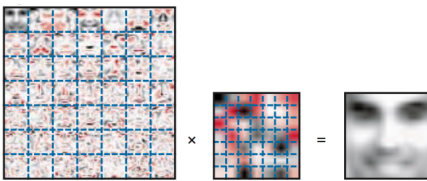


"eigenfaces"

2: VQ

Example (Lee and Seung 1999)

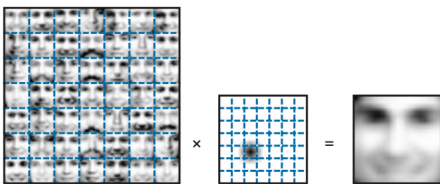
1: PCA



The diagram illustrates the PCA process. On the left is a 10x10 grid of 100 small face images, each with a blue dashed grid overlay. This grid is multiplied (indicated by a 'x' symbol) by a 10x10 weight matrix, which is a small grid of colored squares. The result (indicated by an '=' symbol) is a single grayscale face image. The text "eigenfaces" is written to the right of the resulting image.

\times = "eigenfaces"

2: VQ

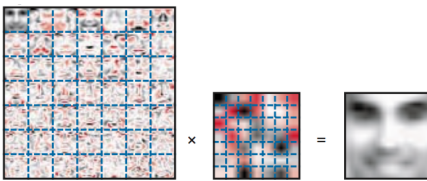


The diagram illustrates the VQ process. On the left is a 10x10 grid of 100 small face images, each with a blue dashed grid overlay. This grid is multiplied (indicated by a 'x' symbol) by a 10x10 weight matrix, which is a small grid of blue squares. The result (indicated by an '=' symbol) is a single grayscale face image.

\times =

Example (Lee and Seung 1999)

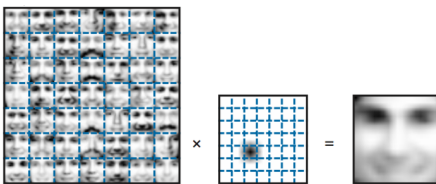
1: PCA



The diagram illustrates the PCA process. On the left is a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by a multiplication symbol 'x' and a 10x10 weight matrix, also with a blue dashed grid. This is followed by an equals sign '=' and a single grayscale face image. The text "eigenfaces" is written to the right of the final image.

\times = "eigenfaces"

2: VQ

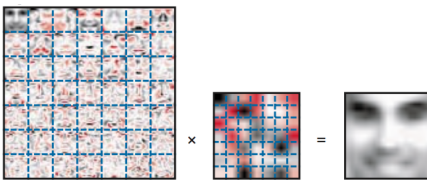


The diagram illustrates the VQ process. On the left is a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by a multiplication symbol 'x' and a 10x10 weight matrix, also with a blue dashed grid. This is followed by an equals sign '=' and a single grayscale face image. The text "protofaces" is written to the right of the final image.

\times = "protofaces"

Example (Lee and Seung 1999)

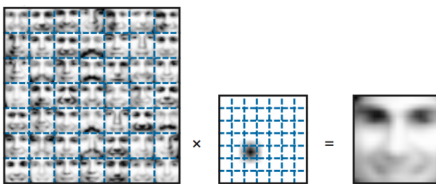
1: PCA



The diagram illustrates the PCA process. On the left is a 10x10 grid of small, noisy face images. To its right is a small 10x10 grid of weights, with some cells highlighted in red. An 'x' symbol is between the grid and the weight grid. To the right of the weight grid is an equals sign, followed by a single grayscale face image labeled "eigenfaces".

$$\text{Grid of face images} \times \text{Weight matrix} = \text{Eigenface}$$

2: VQ



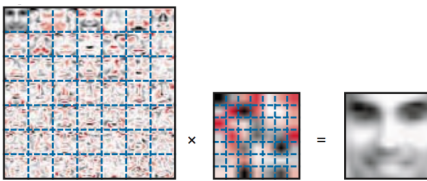
The diagram illustrates the VQ process. On the left is a 10x10 grid of grayscale face images. To its right is a small 10x10 grid of weights, with one cell highlighted in dark blue. An 'x' symbol is between the grid and the weight grid. To the right of the weight grid is an equals sign, followed by a single grayscale face image labeled "protofaces".

$$\text{Grid of face images} \times \text{Prototype matrix} = \text{Protoface}$$

3: NMF

Example (Lee and Seung 1999)

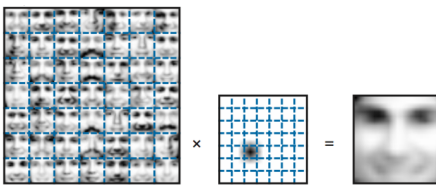
1: PCA



The diagram illustrates the PCA method. On the left is a 10x10 grid of small face images, each with a blue dashed grid overlay. This grid is multiplied (indicated by a 'x' symbol) by a single 10x10 weight matrix, which also has a blue dashed grid. The result (indicated by an '=' symbol) is a single grayscale face image labeled "eigenfaces".

\times $=$ "eigenfaces"

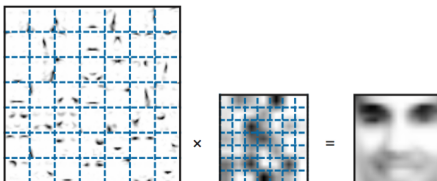
2: VQ



The diagram illustrates the VQ method. On the left is a 10x10 grid of small face images, each with a blue dashed grid overlay. This grid is multiplied (indicated by a 'x' symbol) by a single 10x10 weight matrix, which also has a blue dashed grid. The result (indicated by an '=' symbol) is a single grayscale face image labeled "protofaces".

\times $=$ "protofaces"

3: NMF

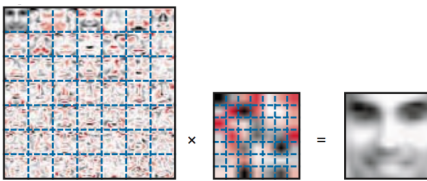


The diagram illustrates the NMF method. On the left is a 10x10 grid of small face images, each with a blue dashed grid overlay. This grid is multiplied (indicated by a 'x' symbol) by a single 10x10 weight matrix, which also has a blue dashed grid. The result (indicated by an '=' symbol) is a single grayscale face image.

\times $=$

Example (Lee and Seung 1999)

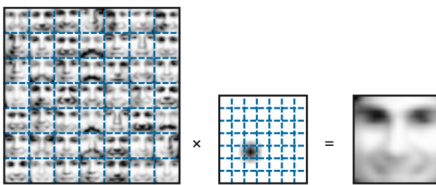
1: PCA



The diagram illustrates the PCA method. On the left is a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by a multiplication symbol (\times), then a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by an equals sign ($=$), then a single grayscale face image.

“eigenfaces”

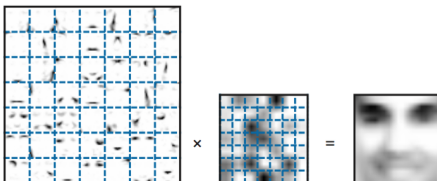
2: VQ



The diagram illustrates the VQ method. On the left is a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by a multiplication symbol (\times), then a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by an equals sign ($=$), then a single grayscale face image.

“protofaces”

3: NMF



The diagram illustrates the NMF method. On the left is a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by a multiplication symbol (\times), then a 10x10 grid of small, noisy face images, each with a blue dashed grid overlay. This is followed by an equals sign ($=$), then a single grayscale face image.

parts-based

NNMF learns an additive, parts based model

NNMF learns an additive, parts based model



NNMF learns an additive, parts based model



which is what our brain does when recognizing images!