Principal Component Analysis
(PCA)

Idea: what does the SVD do, geometrically?

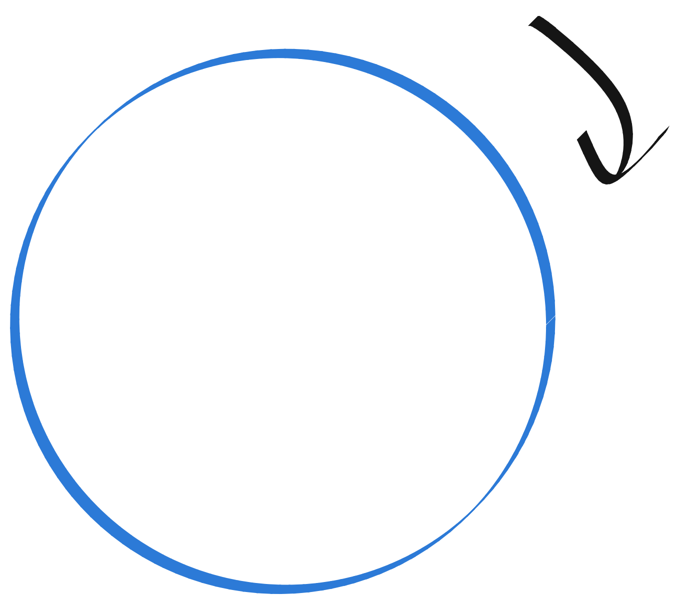
Exploit this understanding to reduce the dimensions of discrete problems

Take Full SVD for A 2×2

$A = U\Sigma V$, U, V are orthogonal 2×2 matrices, Σ is a diagonal 2×2 matrix. Assume 0 is not a diagonal entry of Σ

What does U (or V) do geometrically?

unit circle



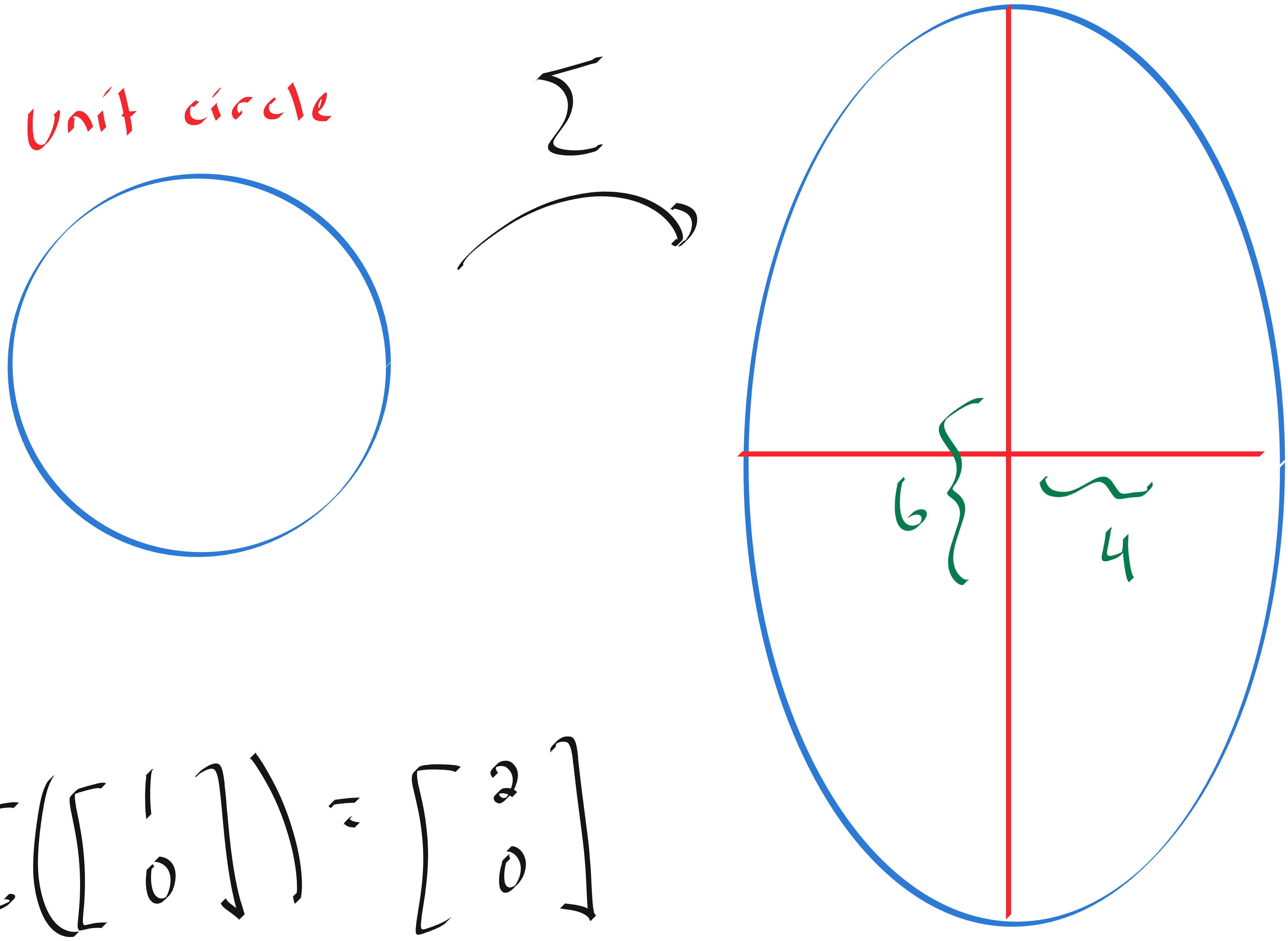
Both U and V rotate the unit circle (can also have reflections)

Diagonal matrices

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Unit circle

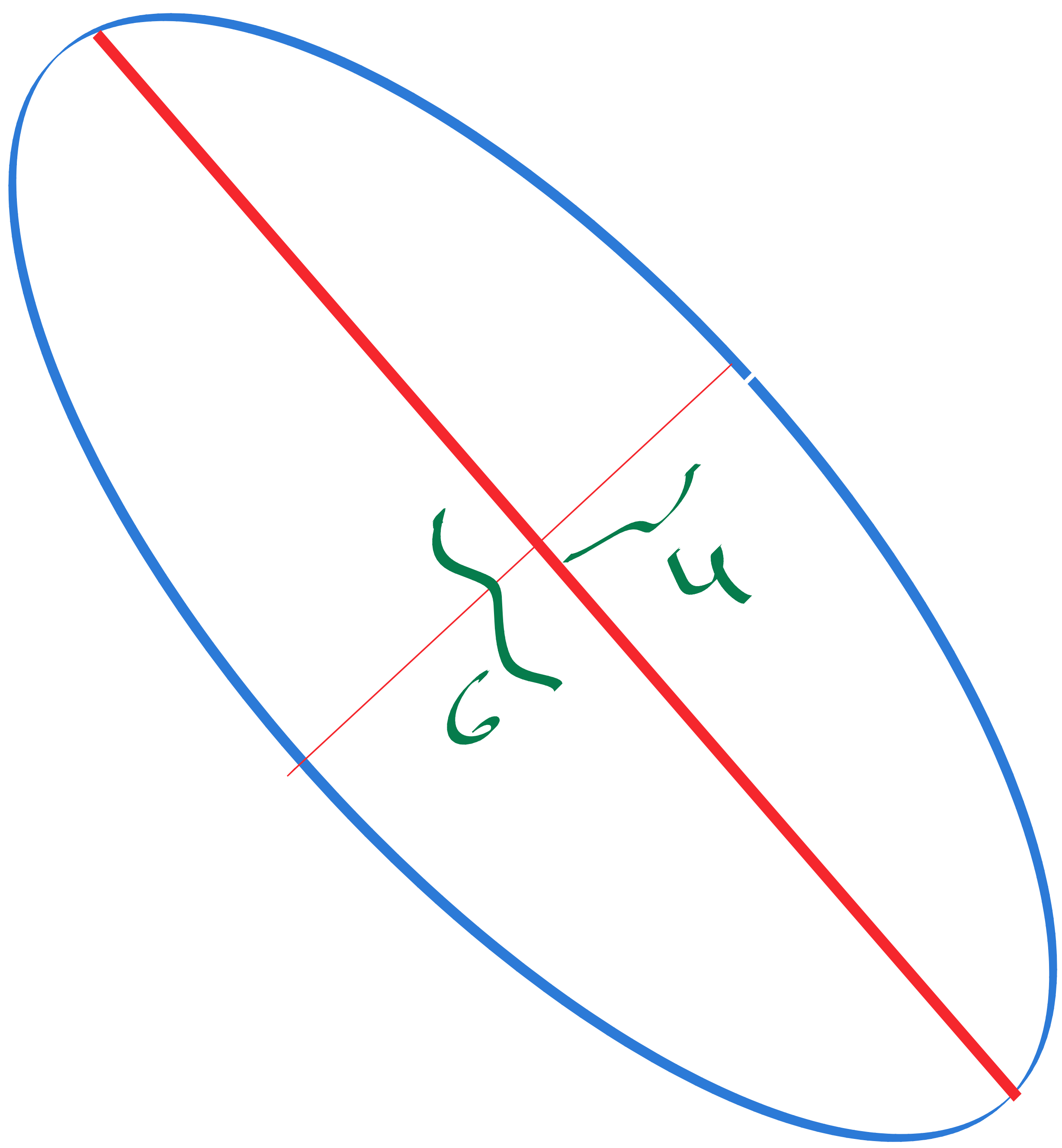
ellipse



$$\Sigma\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\Sigma\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Applying U rotates the ellipse



= A (unit circle)

is an ellipse

Ellipse Centered at (0,0)

Equation: $$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Major and minor axes are associated to either $x=0$ or $y=0$.

Higher Dimensions

Ellipsoids!

Equation in $\mathbb{R}^k$:

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \cdots + \frac{x_k^2}{a_k^2} = 1$$

Axes are associated to $x_i = 0$ for all but one coordinate. This produces line segments, largest one = most relevant information.

Idea of PCA:

1^{st} principal component

= longest axis of the ellipsoid produced by A.

(sometimes called "the" principal component)

2^{nd} principal component

= 2nd longest axis of the ellipsoid produced by A, etc.

Note: by applying permutation matrices, we may assume that, if the diagonal entries of Σ are $\sigma_1, \sigma_2, \ldots, \sigma_k$,

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_k$$

If A is $m \times n$

$A = U \Sigma V$ full SVD

Score matrix: $\boxed{T = U\Sigma}$

1^{st} principal component = 1^{st} column of T

2^{nd} " " = 2^{nd} " " "

etc.

Example 1: $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

Find the principal component.

Solution: Ask Wolfram Alpha!

Wolfram Alpha gives us

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$A = U \Sigma V$$

$$T = U\Sigma$$

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{3} & -1 & 0 \\ \sqrt{3} & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{3/2} & -\frac{1}{\sqrt{2}} & 0 \\ \sqrt{3/2} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$$\boxed{\begin{array}{l} \text{Principal Component} \\ = \begin{bmatrix} \sqrt{3/2} \\ \sqrt{3/2} \end{bmatrix} = \sqrt{3/2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{array}}$$

Application: image recognition

(stolen shamelessly from Section 7.3 in Strang)

photo = digital photo = pixelated image

Every pixel has a numerical value that corresponds to the color of the pixel.

Photos = head shots in 7.3

Suppose each image has the same number of pixels m and suppose you have n images. Take every image of m pixels and make a column vector out of this image ("vectorize" the matrix)

Then "center" each vector by subtracting its 1-norm from each component (averaged over the length of the vector)

Example 2: Start with

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 7 \end{bmatrix}$$

Vectorize!

$$\begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \\ 3 \\ 7 \end{bmatrix} = v$$

$\|v\|_1$ = sum of entries

$= 18$

average: $\dfrac{\|v\|_1}{\text{length}} = \dfrac{18}{6} = 3$

Subtract 3 from every entry of the vector:

$$\begin{bmatrix} -2 \\ -1 \\ 2 \\ -3 \\ 0 \\ 4 \end{bmatrix}$$

Sum of entries = 0 ✓

this is a "centered" vector.

Centered = mean is zero

Now take the vectors you have obtained through vectorizing and averaging each image, and arrange them as the columns of a matrix A. The 1st principal component of A gives the linear combination of known images that best identifies a new image. The 2nd principal component gives the second-best identification, etc.

Example 3:

Start with images

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 7 \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} 12 & 0 & 8 \\ 4 & 10 & 14 \end{bmatrix}$$

We've already vectorized and centered the first matrix, so let's do the second:

vectorize as $\begin{bmatrix} 12 \\ 0 \\ 8 \\ 4 \\ 10 \\ 14 \end{bmatrix} = w$

$\|w\|_1 = 48$

average: $\frac{48}{6} = 8$

Center w:

$$\begin{bmatrix} 4 \\ -8 \\ 0 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

Sum of entries is zero.

$$A = \begin{bmatrix} -2 & 4 \\ -1 & -8 \\ 2 & 0 \\ -3 & -4 \\ 0 & 2 \\ 4 & 6 \end{bmatrix}$$

SVD is rather complicated!

Problem: Size!

each image is, say,

640 x 480 pixels.

Each vector is then in

$\mathbb{R}^{307,200}$.

Suppose you have 1,000,000 such images. This is a lot of data! Maybe too much for your computer.

Fix: compute Σ ahead of time.

Σ will have 307,200 diagonal entries. Only take a fixed number of these entries, say the first 500 singular values.

Use this to generate smaller matrices with enough of the "important" data.

is 500 enough? Is it too much?

Trick for images: most nearby pixels share the same color value. Average certain portions of your image to produce a smaller matrix.