Singluar Value Decomposition (SVD)

Dávid Iván

June 16, 2022

Contents

1 SVD without proofs

Let X be an n-by-m matrix, with rank $r \leq \min(m, n)$. Then we can write X as the sum of r one-rank matrices:

$$X = \sum_{i=1}^{r} \sigma_i (u_i v_i^T) \tag{1}$$

where σ_i is a singular value, always positive, and we assume they are ordered, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

The u-vectors form an orthonormal basis in the column space $R(X)$. The v-vectors form an orthonormal basis in the row space $R(X^T)$.

2 SVD of a centered data matrix

Let X be an N-by-p data matrix, where each row is a data point. We center the data matrix, so that it has zero mean (the sum of each column is zero).

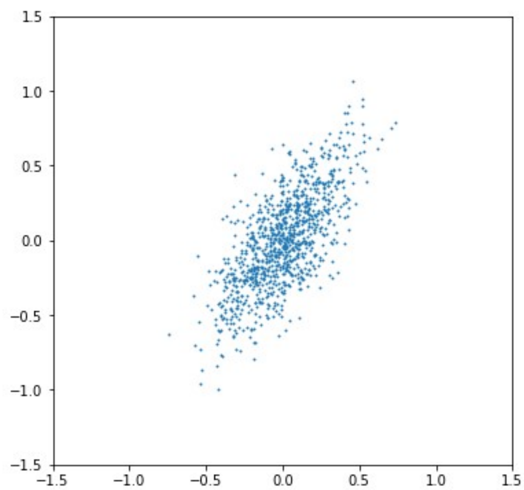

Figure 1: Data. The rows of X are thought of as data points.

Consider a unit vector (in the row space) a. We can project the data points onto W_a (the line defined by a) with the projection matrix:

$$P_a = \frac{aa^T}{a^T a} = aa^T \tag{2}$$

The projection of v onto W_a:

$$v' = P_a v = aa^T v = (a^T v)a \tag{3}$$

We see that the projected point is on the line of a, and its coordinate (distance from origin) is $a^T v$ (or $v^T a$). We can project all data points, the coordinates are Xa. The variance of the data points along the line:

$$\text{Var}(Xa) = \frac{1}{N}||Xa||^2 \tag{4}$$

From this we see that if we want to maximize the variance (finding the first principal direction), a must be v_1, and then the variance:

$$\text{Var}(Xv_1) = \frac{1}{N}||Xv_1||^2 = \frac{1}{N}||\sigma_1 u_1||^2 = \frac{\sigma_1^2}{N} \tag{5}$$

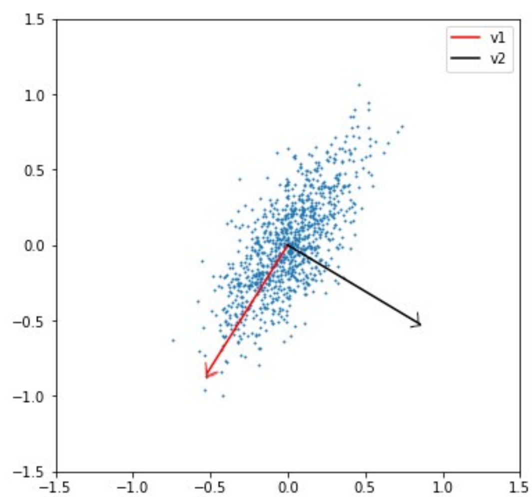

Figure 2: SVD of the data matrix. v_1 is pointing to the largest variance, v_2 is orthogonal to it.

So the diagonal form of the covariance matrix is D^2/N. The sample covariance:

$$S = \frac{1}{N}X^T X = \frac{1}{N}VD^2V^T = V\frac{D^2}{N}V^T \tag{6}$$

3 Image compression

Considering an image as a matrix A (H-by-W), the SVD gives a possibility to compress the image by considering the most relevant contributions, i.e. those that have the biggest singular values.

$$A = \sum_{i=1}^{r} \sigma_i \cdot u_i \cdot v_i^T \quad \rightarrow \quad \hat{A} = \sum_{i=1}^{c} \sigma_i \cdot u_i \cdot v_i^T \tag{7}$$

where $c << r$ is the number of components to consider in the new sum. The following figure shows an example of how it performs.

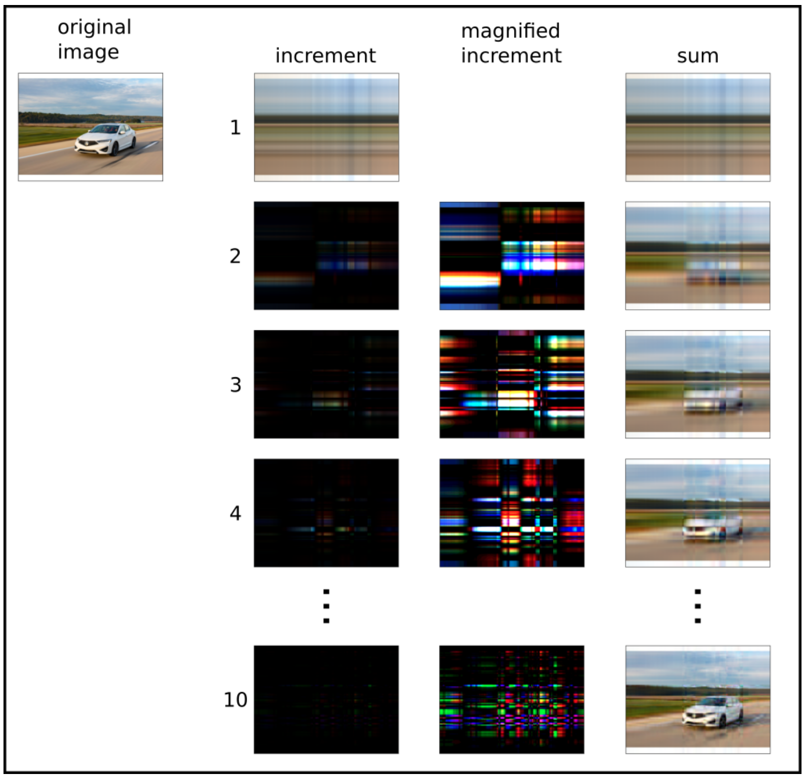

Figure 3: Image compression with SVD. Color channels are treated as independent matrices. Here the rank of the (let's say the red) matrix is $r \approx 656$. Taking the first 10 components gives a pretty nice result.