Computer Science 414 Notes

Joaquin Schulze

February 2020

1 PCA (Principle Component Analysis)

1.1 Idea

By using PCA we can reduce the dimensions of data and therefore create a more manageable and meaningful data structure by choosing new lower dimensional sub-spaces that contain the essential data of the larger subspace.

For instance a black and white image contains many shades of grey, this can be reduced to a lower dimensional subspace that only contains greys that are a certain degree of darker or lighter than the previous grey.
Note: Just because we say that it chooses a lower subspace does not mean it discards the old data, it simply means it is generating a new characteristic from the old ones that summarises the data in a lower subspace.

1.2 Mathematical methodology

1.2.1 Step 1 : Subtract the mean of the data

PCA requires the data set to have a mean of zero, therefore we subtract the mean of each dimension with the respective data from that dimension.

1.2.2 Step 2 : Calculate the co variance matrix

If we work with a 2 dimensional data set we will have a covariance matrix of 2x2.
Notice: If the off diagonal elements are positive then the correlation between the data is implied to be positive (The data in the two dimensions will increase together)

1.2.3 Step 3 : Get the eigenvalue decomposition of the covariance matrix

If the covariance matrix is square and symmetric we can get a normalised eigenvector matrix and a diagonal eigenvalue matrix.
Notice: Both eigenvectors are unitary vectors

Notice: Like expected both eigenvectors are perpendicular to each other
The direction of the eigenvectors of the largest eigenvalue show the strongest data pattern in the data set.

1.2.4 Step 4: Choose the significant principle components

As we saw in step 3, after ordering our eigenvalues we choose the largest one to make sure we have the smallest loss margin of data.

1.2.5 Step 4: Get the transformed data

We now need to convert the data into one dimension as it us still in 2 dimensions after step 4. We then need to project the data onto the new eigenvector and then rotate the vector to one dimensional space (line).
Note: this can be done in one step by multiplying the original data with the feature vector

1.3 Mathematical Derivation

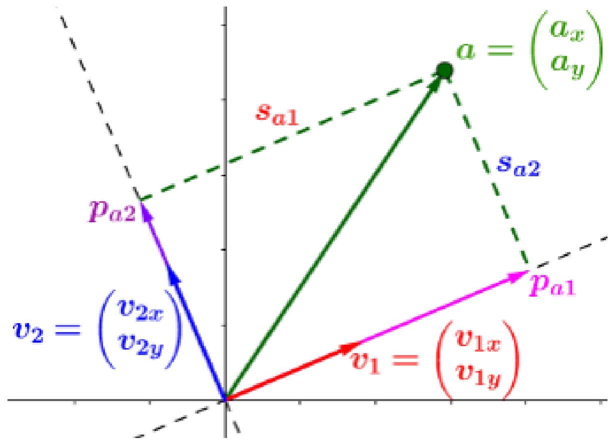
I dont have enough time to convert this into Latex for assignment one, but https://www.projectrhea.org/rhea/index.php/PCA has a good explanation at the bottom of the page if the class notes confuses you

2 SVD (Singular Value Decomposition)

2.1 Idea

The idea is to split a data point on an axis into it's projection direction unit vector and the lengths of the projection onto it.

As shown below we can get the lengths of the data point from the direction vector (usually close to the principle direction vector , refer to section on PCA's, called the decomposition axis)



Same figure as before, but tilting the axes of projection to convince you they aren't confined to x and y. (a_x and a_y are the coordinates of vector a, put into a column matrix (aka column vector), as per convention. Same for v_1 and v_2).

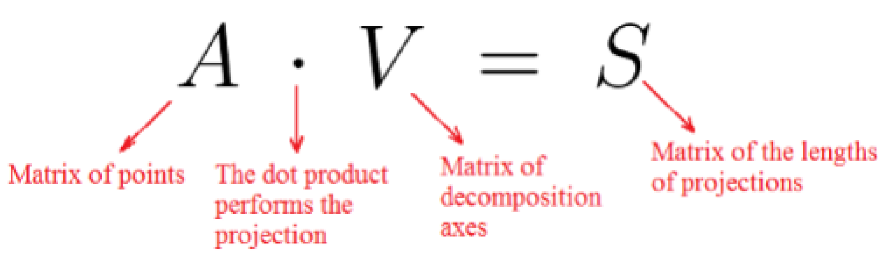
We want to decompose (project) the vector $\boldsymbol{a}$ along unit vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$.

2.2 Mathematical methodology

SVD builds on PCA's and is just a method of representing the data on the new principle/decomposition axis

2.2.1 Step 1 : Project the data onto the decomposition axis Via dot product

If we take the dot product of the data with the vector of the principle axis we will get the lengths of each point from the decomposition axis.



2.2.2 Step 2: Isolate A

We want A to be alone so that we can express it in terms of other matrices, because the matrix of V contains orthonormal columns, it's inverse is it's transpose. Thus we now have

$$\begin{aligned} A \cdot V &= S \\ A &= S \cdot V^T \end{aligned} \tag{1}$$

2.3 Step 3: We want to split the length matrix (Magnitude matrix) up into its normalised matrices

Notice: We now call the S matrix the magnitude matrix and denote the sign as M

We want to split the magnitude matrix into a singular value matrix that is the diagonal of the magnitude of each column of the magnitude matrix (denoted as the σ matrix) and its orthogonal matrix (denoted as U).

To normalize the columns of S, we divide them by their magnitude…

$$S = \begin{pmatrix} s_{a1} & s_{a2} \\ s_{b1} & s_{b2} \end{pmatrix}$$

$$\text{Magnitude of 1st column} = \sigma_1 = \sqrt{(s_{a1})^2 + (s_{b1})^2}$$

$$\text{Magnitude of 2nd column} = \sigma_2 = \sqrt{(s_{a2})^2 + (s_{b2})^2}$$

…by doing with S what we did with M in the example above:

$$S = \begin{pmatrix} \frac{s_{a1}}{\sigma_1} & \frac{s_{a2}}{\sigma_2} \\ \frac{s_{b1}}{\sigma_1} & \frac{s_{b2}}{\sigma_2} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \underbrace{\begin{pmatrix} u_{a1} & u_{a2} \\ u_{b1} & u_{b2} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}}_{\Sigma}$$

Finally…

$$\underbrace{A = U \, \Sigma \, V^T}_{\text{Singular Value Decomposition}}$$

2.3.1 Mathematical Derivation

Find in class notes on page 69 to 73