CS 663 : Digital Image Processing : Assignment 4

Siddharth Saha [170100025], Tezan Sahu [170100035]

Due Date :- 16rd October, 2019

Question 5

Question:
Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$, with average vector $\bar{\boldsymbol{x}}$. We have seen in class that the direction $\boldsymbol{e}$ such that $\sum_{i=1}^{N} \|\boldsymbol{x_i} - \bar{\boldsymbol{x}} - (\boldsymbol{e} \cdot (\boldsymbol{x_i} - \bar{\boldsymbol{x}}))\boldsymbol{e}\|^2$ is minimized, is obtained by maximizing $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{e}$, where $\boldsymbol{C}$ is the covariance matrix of the vectors in $\mathcal{X}$. This vector $\boldsymbol{e}$ is the eigenvector of matrix $\boldsymbol{C}$ with the highest eigenvalue. Prove that the direction $\boldsymbol{f}$ perpendicular to $\boldsymbol{e}$ for which $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$ is maximized, is the eigenvector of $\boldsymbol{C}$ with the second highest eigenvalue. For simplicity, assume that all non-zero eigenvalues of $\boldsymbol{C}$ are distinct and that $\operatorname{rank}(\boldsymbol{C}) > 2$.

Solution:
We use Lagrange Multipliers to solve this constrained optimization problem. Let:

$$F(\boldsymbol{f}) = \boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f} - \lambda_1(\boldsymbol{f}^t\boldsymbol{f} - 1) - \lambda_2(\boldsymbol{f}^t\boldsymbol{e})$$

Taking gradient of $F(\boldsymbol{f})$ wrt $\boldsymbol{f}$ and setting it to 0 to find the optima, we get:

$$\boldsymbol{C}\boldsymbol{f} - \lambda_1\boldsymbol{f} - \lambda_2\boldsymbol{e} = 0 \tag{1}$$

$\boldsymbol{e}$ is a unit vector. Thus, pre-multiplying by $\boldsymbol{e}^t$, we get:

$$\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} - \lambda_1\boldsymbol{e}^t\boldsymbol{f} - \lambda_2\boldsymbol{e}^t\boldsymbol{e} = 0$$

$$\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} - \lambda_1\boldsymbol{e}^t\boldsymbol{f} - \lambda_2 = 0 \tag{2}$$

Now, $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} = (\boldsymbol{C}^t\boldsymbol{e})^t\boldsymbol{f}$
But we know that $\boldsymbol{C}$ is symmetric, ie, $\boldsymbol{C}^t = \boldsymbol{C}$
Thus, we have:

$$\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} = (\boldsymbol{C}\boldsymbol{e})^t\boldsymbol{f} = \mu_1\boldsymbol{e}^t\boldsymbol{f} \tag{3}$$

Here, μ_1 is the largest eigenvalue of $\boldsymbol{C}$, corresponding to eigenvector $\boldsymbol{e}$ We also said that $\boldsymbol{e}$ is perpendicular to $\boldsymbol{f}$, ie, $\boldsymbol{e}^t\boldsymbol{f} = 0$. Substituting this & (3) in (2), we get:

$$(\mu_1 - \lambda_1)\boldsymbol{e}^t\boldsymbol{f} - \lambda_2 = 0$$

$$\therefore \lambda_2 = 0 \tag{4}$$

Substituting (4) in (1), we get:

$$\boldsymbol{C}\boldsymbol{f} = \lambda_1\boldsymbol{f}$$

This clearly shows that $\boldsymbol{f}$ is an eigenvector of $\boldsymbol{C}$, with eigenvalue λ_1. Now, $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f} = \lambda_1$ and $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$ had to be maximized. It is also given that non-zero eigenvalues of $\boldsymbol{C}$ are distinct.
Thus, λ_1 must be the second largest eigenvalue of $\boldsymbol{C}$