CS 663 Assignment 4: PCA and SVD

Question 5

October 12, 2018

Proof goes as follows:

Our Objective function for $\boldsymbol{f}$ that needs to be maximized is as follows:
$J(\boldsymbol{f}) = \boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$

As $\boldsymbol{f}$ is a unit vector that denotes direction perpendicular to $\boldsymbol{e}$, there are two constraints on $\boldsymbol{f}$ which are as follows:
$\|\boldsymbol{f}\|^2 = \boldsymbol{f}^t\boldsymbol{f} = 1$
$\boldsymbol{e}\cdot\boldsymbol{f} = \boldsymbol{e}^t\boldsymbol{f} = \boldsymbol{f}^t\boldsymbol{e} = 0$

We will use the concept of Lagrange multiplier to maximize our objective function with respect to the above two constraints.

The Lagrange function will be as follows:

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

Taking derivative of $\widetilde{J}(\boldsymbol{f})$ with respect to $\boldsymbol{f}$ and setting it to 0, we get:
$2\boldsymbol{C}\boldsymbol{f} - 2\lambda_1\boldsymbol{f} - \lambda_2\boldsymbol{e} = 0$

Multiplying this equation by $\boldsymbol{e}^t$, we get:
$2\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} - 2\boldsymbol{e}^t\lambda_1\boldsymbol{f} - \lambda_2\boldsymbol{e}^t\boldsymbol{e} = 0$
As C is symmetric, we can write the equation as:
$2\boldsymbol{e}^t\boldsymbol{C}^t\boldsymbol{f} - 2\lambda_1(\boldsymbol{e}^t\boldsymbol{f}) - \lambda_2(\boldsymbol{e}^t\boldsymbol{e}) = 0$
$\implies 2(\boldsymbol{C}\boldsymbol{e})^t\boldsymbol{f} - 2\lambda_1(\boldsymbol{e}^t\boldsymbol{f}) - \lambda_2(\boldsymbol{e}^t\boldsymbol{e}) = 0$
We know that $\boldsymbol{e}$ is an eigenvector of C, so $\boldsymbol{C}\boldsymbol{e} = \lambda_e\boldsymbol{e}$ where λ_e is the eigenvalue associated with the eigenvector $\boldsymbol{e}$. With this information, we get:
$2(\lambda_e\boldsymbol{e})^t\boldsymbol{f} - 2\lambda_1(\boldsymbol{e}^t\boldsymbol{f}) - \lambda_2(\boldsymbol{e}^t\boldsymbol{e}) = 0$
$\implies 2\lambda_e(\boldsymbol{e}^t\boldsymbol{f}) - 2\lambda_1(\boldsymbol{e}^t\boldsymbol{f}) - \lambda_2(\boldsymbol{e}^t\boldsymbol{e}) = 0$
As $\boldsymbol{e}$ and $\boldsymbol{f}$ are perpendicular and $\boldsymbol{e}$ is a unit vector, we get:
$\lambda_2 = 0$

So our derivative of $\widetilde{J}(\boldsymbol{f})$ with respect to $\boldsymbol{f}$ becomes: $2\boldsymbol{C}\boldsymbol{f} - 2\lambda_1\boldsymbol{f} = 0$
$\implies \boldsymbol{C}\boldsymbol{f} = \lambda_1\boldsymbol{f}$
Hence, we have proven that $\boldsymbol{f}$ is an eigen vector of the covariance matrix $\boldsymbol{C}$ (with eigenvalue as λ_1). As $\boldsymbol{C}$ is a symmetric matrix, this is also a convenient result as eigenvectors of a symmetric matrix are orthogonal and our second constraint enforced the orthogonality of $\boldsymbol{e}$ and $\boldsymbol{f}$.

With the result obtained from lagrange multiplier, we can rewrite our objective function as following:
$J(\boldsymbol{f}) = \boldsymbol{f}^t\lambda_1\boldsymbol{f}$
$\implies J(\boldsymbol{f}) = \lambda_1\boldsymbol{f}^t\boldsymbol{f}$
As $\boldsymbol{f}$ is a unit vector:

$J(\boldsymbol{f}) = \lambda_1$

Now we want to maximize our objective function, so we want λ_1 to be the highest possible eigenvalue. But $\boldsymbol{e}$ is the eigenvector of $\boldsymbol{C}$ corresponding to the highest eigenvalue and from the orthogonality of $\boldsymbol{e}$ and $\boldsymbol{f}$, we can infer that $\boldsymbol{e} \neq \boldsymbol{f}$.
Therefore, λ_1 is the second highest eigenvalue of the covariance matrix $\boldsymbol{C}$ and $\boldsymbol{f}$ is the eigenvector of $\boldsymbol{C}$ corresponding to the second highest eigenvalue.