**Machine Learning**

**CSCI 567 Spring 2021**

Discussion: PCA

Date: Apr 2th, 2021

In this problem, we use proof by induction to show that the $M$-th principle component corresponds to the $M$-th eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$ sorted by the eigenvalue from largest to smallest. Here $\boldsymbol{X}$ is the centered data matrix and we denote the sorted eigenvalues as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$. In the lecture, the results was proven for $M = 1$. Now suppose the result holds for a value $M$, and you are going to show that it holds for $M + 1$. Note that the $M + 1$ principle component corresponds to the solution of the following optimization problem:

$$\max_{\boldsymbol{v}} \quad \boldsymbol{v}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{v} \tag{1}$$

$$\text{s.t.} \quad \|\boldsymbol{v}\|_2 = 1 \tag{2}$$

$$\boldsymbol{v}^T\boldsymbol{v}_i = 0, i = 1, ..., M \tag{3}$$

where $\boldsymbol{v}_i$ is the $i$-th principle component. Write down the Lagrangian of the optimization problem above, and show that the solution $\boldsymbol{v}_{M+1}$ is an eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$. Then show that the quantity in (1) is maximized when the $\boldsymbol{v}_{M+1}$ is the eigenvector with eigenvalue $\lambda_{M+1}$.