PCA quiz

CMSC498 Spring 2015

April 21, 2015

Name(s):
UID(s):

Suppose you are given a dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_i$ is the vector of p predictor values for the i-th observation.

I just ran the Principal Components Algorithm to obtain linear transformations

$$[\phi_1, \phi_2, \ldots \phi_p]$$

And have calculated matrix

$$Z_{n\times 2} = X_{n\times p}\,[\phi_1, \phi_2]_{p\times 2}$$

and now want to analyze the results: i.e., vectors ϕ_j and matrix Z.

1) What is the length of each vector ϕ_j?

2) What would I include in a scatter plot showing how observations are embedded in the two-dimensional space defined by the first two principal components of $\mathbf{X}$?

3) What would I include in a scatter plot showing how predictors are correlated based on the first two principal components of $\mathbf{X}$.

4) In principal components analysis, should I always scale the columns of $\mathbf{X}$ to have equal variance? Why?