The University of Texas at Austin
Homework Assignment 8

Predictive Analytics

October 23, 2026

Instructions: Provide your complete solution to the following problems. Final answers only, without appropriate justification, will receive zero points even if correct.

Principal Components Analysis.

Problem 8.1. (15 points) Before computing principal components, the variables are centered to have mean zero. What would the first loading vector tend to capture if the data were **not** centered?

Problem 8.2. ($5 + 5 + 5 = 15$ points) An analyst runs PCA on 20 standardized rating variables and makes the following claims. For each, state whether you agree/disagree and explain your reasoning.

i. *"PC1 explains only* 30% *of the variance, so PCA was not useful."*

ii. *"Variable* A *has a PC1 loading of* 0.02*, so* A *is unimportant for predicting claim costs and can be dropped from my model."*

iii. *"The PC scores are uncorrelated, so using them as predictors in a linear regression avoids multicollinearity."*

Problem 8.3. ($5 + 5 + 5 + 5 = 20$ points) Two **standardized** variables X_1 and X_2 have correlation r. The principal component loading vectors are

$$\varphi_1 = \frac{1}{\sqrt{2}}(1, 1), \quad \varphi_2 = \frac{1}{\sqrt{2}}(1, -1),$$

with score variances $1 + r$ and $1 - r$, respectively.

i. For $r = 0.8$, compute the PVE of each component.
ii. If it were $r = -0.8$, which loading vector would define the first principal component? What would its PVE be?
iii. What happens when $r = 0$? Is PCA useful for dimension reduction in this case?
iv. Using your answers, explain in one or two sentences when PCA is at its most effective.