Michigan State University **March 15, 2024**
CMSE381 - Data Science

CMSE381 - Quiz 7

I will adhere to the Spartan Code of Honor in completing this assignment.

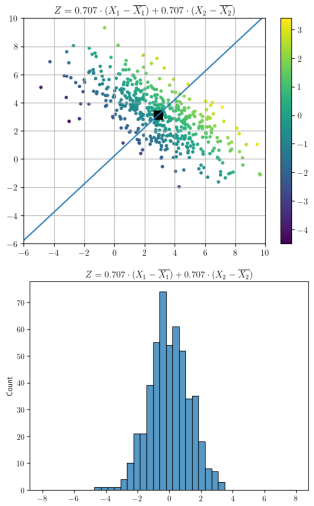
Signed: _____

1. (3pts) In doing dimension reduction such as PCA, we have input variables $X_1, \cdots, X_p$ and we want to construct new predictors $Z_1, \cdots, Z_M$ that are linear combinations of the X_i's. What are we aiming for in terms of M and p?

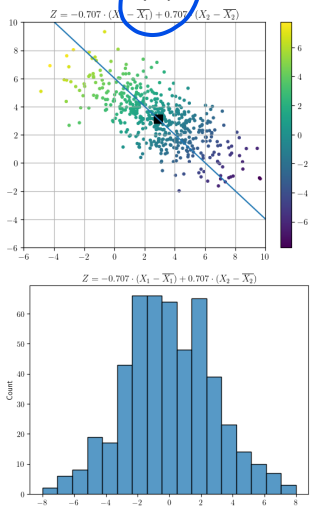
   A. $p < M$ B. $p = M$ C. $M < p$

2. (3pts) The following figures give the projection information for several cases of data with $p = 2$ (so the axis labels should be X_1 and X_2 for each), along with the distribution of resulting Z_1 values for each. Circle the letter for one which is closest to the line that would be found as the 1st principal component (1st PC).

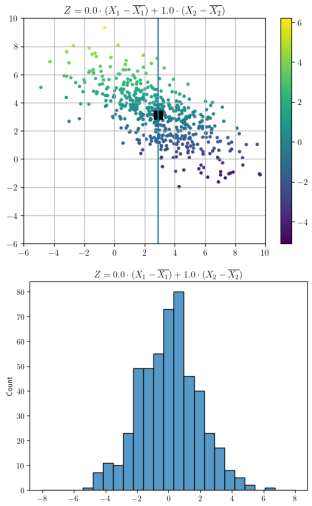
(A)



(B)



(C)



3. (4pts) This question tests your high-level understanding of the algorithms, your answer does not need to be very accurate.
   (a) If we apply Lasso to the dataset given in the table with appropriate choice of λ found by the Cross Validation, which coefficient is more likely to become 0? Why?
   (b) If we apply PCR to this dataset and set $M = 2$, what direction is likely to be discarded? Why?

y	X1	X2	X3
1	1	2	1
3	2	-1	1
4	1	-2	1

(a) β_3, since X_3 has no information (all values of X_3 are the same)

(b) $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, which is the direction along X_3 since this direction has the minimal var