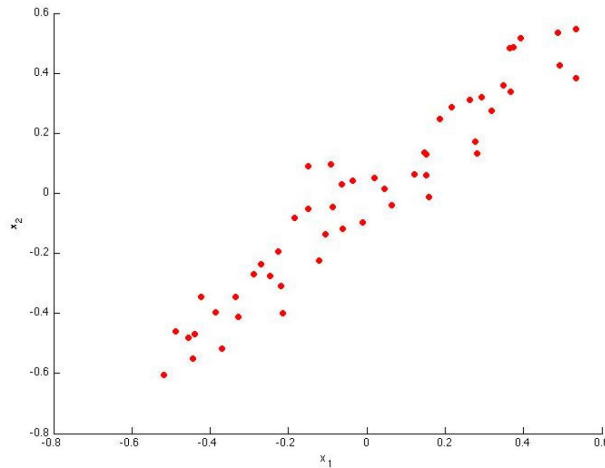


HMC CS 158

Quiz 8: PCA

1. Consider applying PCA to the following 2D dataset:



On the figure, draw the (two) vectors that could correspond $\mathbf{u}^{(1)}$ (the first eigenvector / first principal component).

2. Which of the following is a reasonable way to select the number of principal components k ?
- (a) Choose the appropriate answers: Choose k to be the largest / smallest value so that at least 99% / 1% of the variance is retained.
 - (b) True or False: We can use the elbow method to select k .
3. Which of the following statements are true? Check all that apply.
- (a) If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA.
 - (b) Even if all the input features are on very similar scales, we should still perform mean normalization (so that each feature has zero mean) before running PCA.
 - (c) Feature scaling is not useful for PCA, since the eigenvector calculation (such as using Python's `numpy.linalg.eig`) takes care of this automatically.
 - (d) PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).

4. Which of the following are recommended applications of PCA? Select all that apply.
- (a) Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space.
 - (b) Data visualization: Reduce data to 2D (or 3D) so that it can be plotted.
 - (c) Data visualization: To take 2D data and find a different way of plotting it in 2D (using $k = 2$).
 - (d) As a replacement for (or alternative to) linear regression. For most learning applications, PCA and linear regression give substantially similar results.