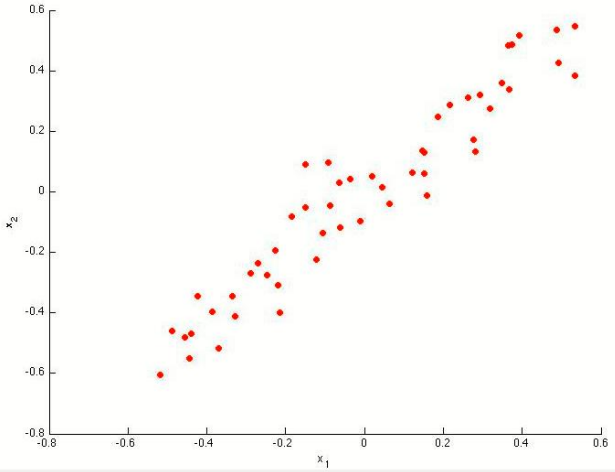
Principal Component Analysis

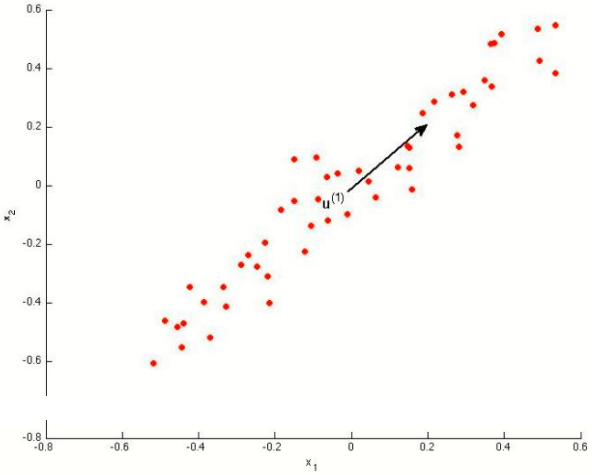
TOTAL POINTS 5

1. Consider the following 2D dataset: 1 point

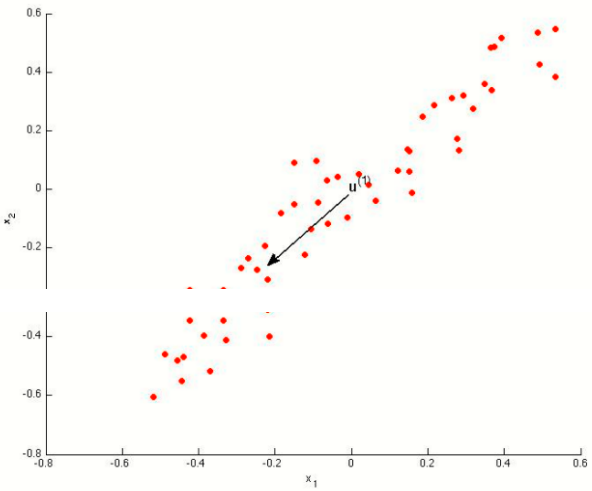


Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).

- [x]



- [x]



2. Which of the following is a reasonable way to select the number of principal components k? 1 point

(Recall that n is the dimensionality of the input data and m is the number of input examples.)

- (●) Choose k to be the largest value so that at least 99% of the variance is retained
- ( ) Use the elbow method.
- ( ) Choose k to be 99% of m (i.e., $k = 0.99 * m$, rounded to the nearest integer).
- ( ) Choose k to be the smallest value so that at least 99% of the variance is retained.

3. Suppose someone tells you that they ran PCA in such a way that "95% of the variance was retained." What is an equivalent statement to this? 1 point

- ( ) $\frac{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}-x_{\text{approx}}^{(i)}||^2}{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}||^2} \geq 0.95$
- ( ) $\frac{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}||^2}{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}-x_{\text{approx}}^{(i)}||^2} \geq 0.95$
- (●) $\frac{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}-x_{\text{approx}}^{(i)}||^2}{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}||^2} \leq 0.05$
- ( ) $\frac{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}-x_{\text{approx}}^{(i)}||^2}{\frac{1}{m}\sum_{i=1}^{m}||x^{(i)}||^2} \geq 0.05$

2 Question: Will be the option 4 (smallest value)

4. Which of the following statements are true? Check all that apply. 1 point

- [x] Given an input $x \in \mathbb{R}^n$, PCA compresses it to a lower-dimensional vector $z \in \mathbb{R}^k$.
- [ ] PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).
- [ ] Feature scaling is not useful for PCA, since the eigenvector calculation (such as using Octave's `svd(Sigma)` routine) takes care of this automatically.
- [x] If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA.

5. Which of the following are recommended applications of PCA? Select all that apply. 1 point

- [x] Data visualization: Reduce data to 2D (or 3D) so that it can be plotted.
- [x] Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space.
- [ ] Preventing overfitting: Reduce the number of features (in a supervised learning problem), so that there are fewer parameters to learn.
- [ ] To get more features to feed into a learning algorithm.