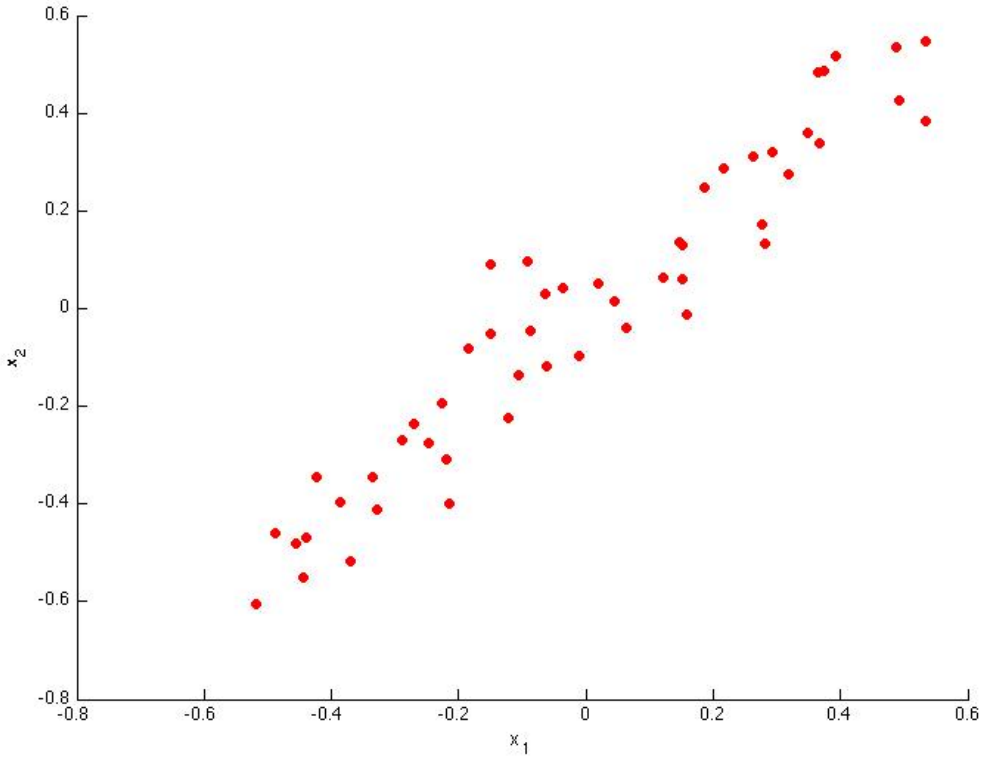
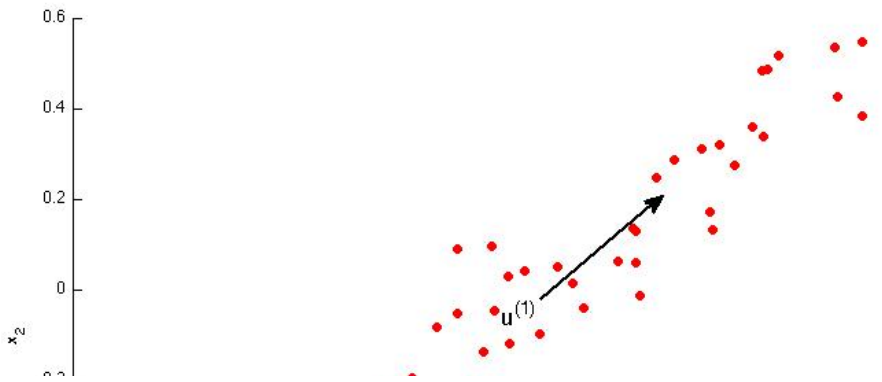
Principal Component Analysis

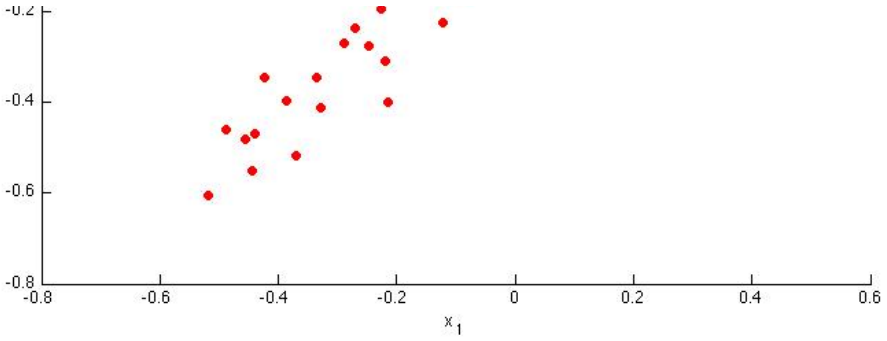
TOTAL POINTS 5

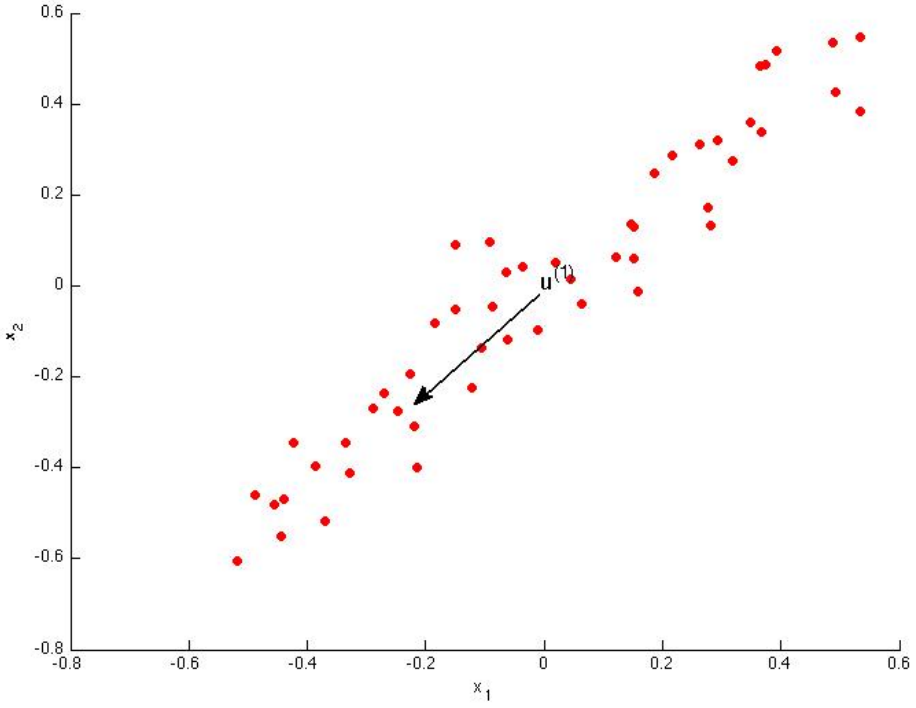
1. Consider the following 2D dataset: 1 point

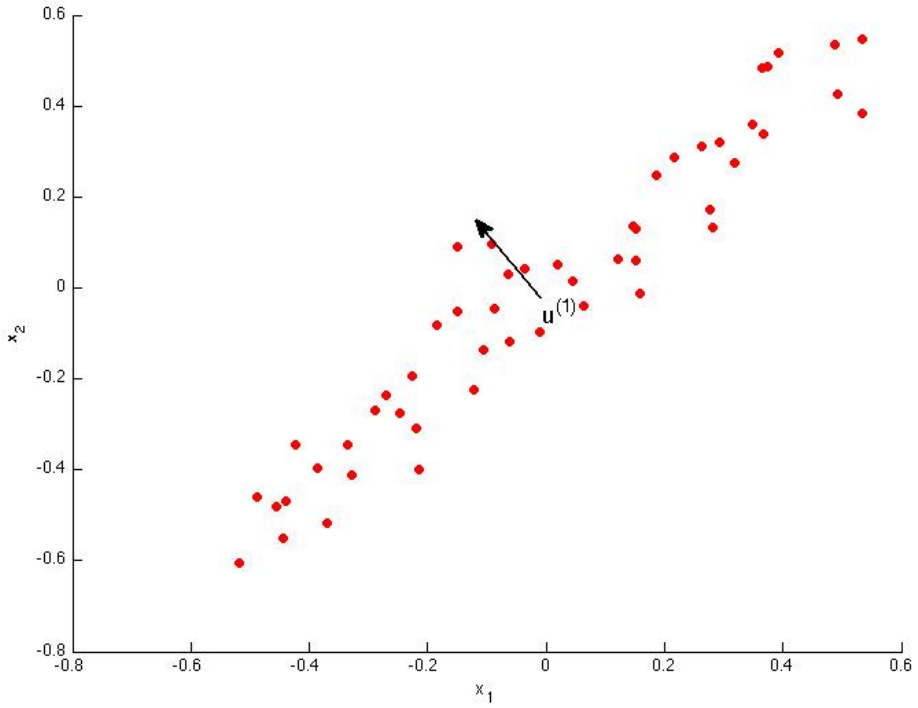


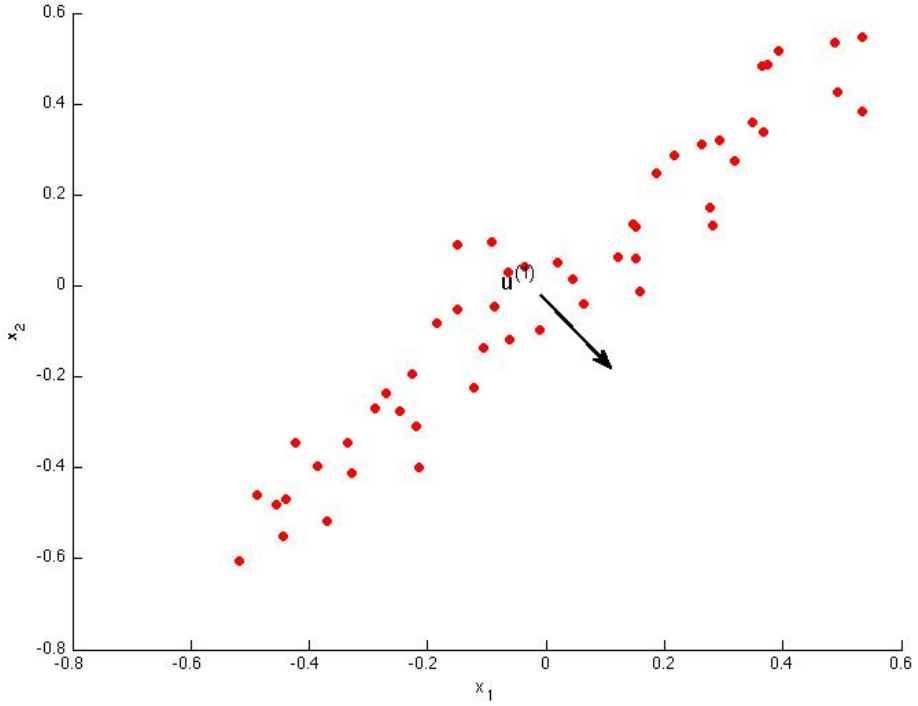
Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).




-0.2
-0.4
-0.6
-0.8
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
x1


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
x2
u(1)
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
x1


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
x2
x1
u(1)



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.)

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

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

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

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

- [x] Given an input $x \in \mathrm{R}^n$, PCA compresses it to a lower-dimensional vector $z \in \mathrm{R}^k$.
- [ ] Feature scaling is not useful for PCA, since the eigenvector calculation (such as using Octave's $\mathrm{svd}(\mathrm{Sigma})$ routine) takes care of this automatically.
- [ ] PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).
- [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 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.
- [x] Data visualization: Reduce data to 2D (or 3D) so that it can be plotted.
- [ ] To get more features to feed into a learning algorithm.

I, **Hassan Rasheed**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera's Honor Code