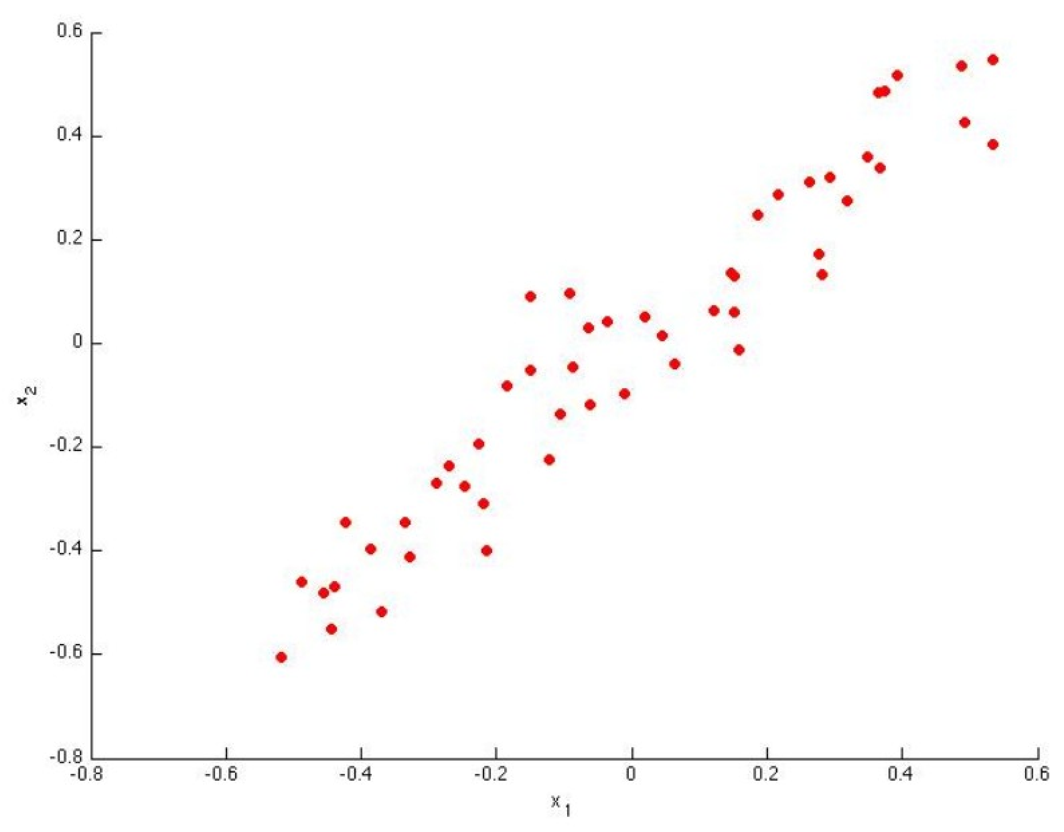
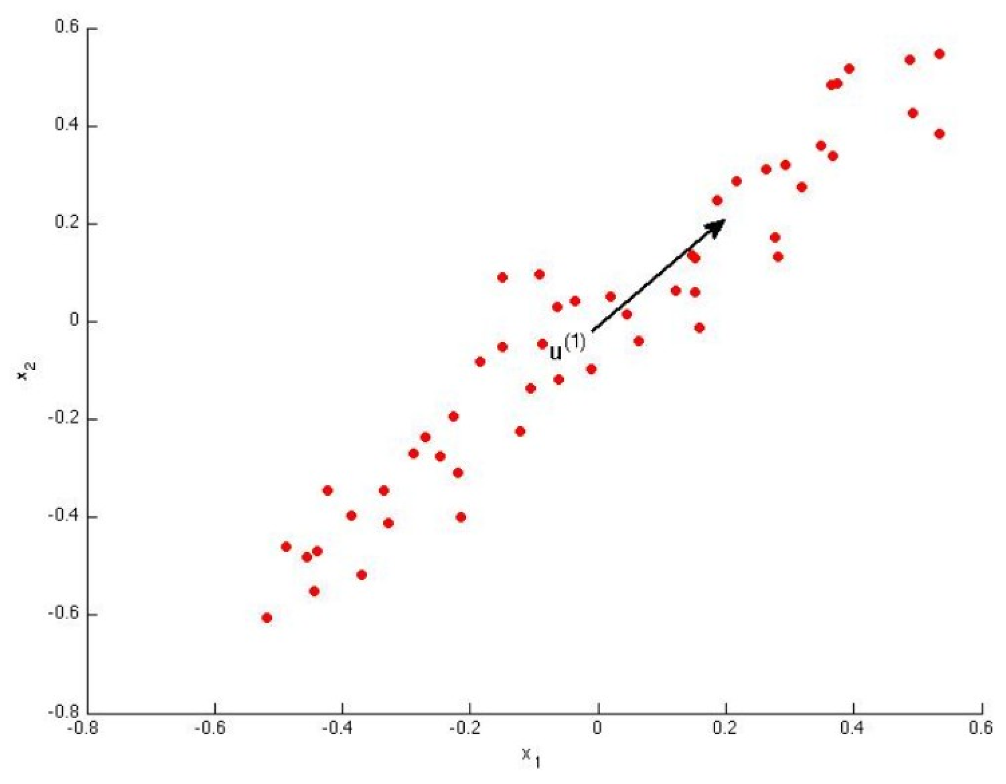
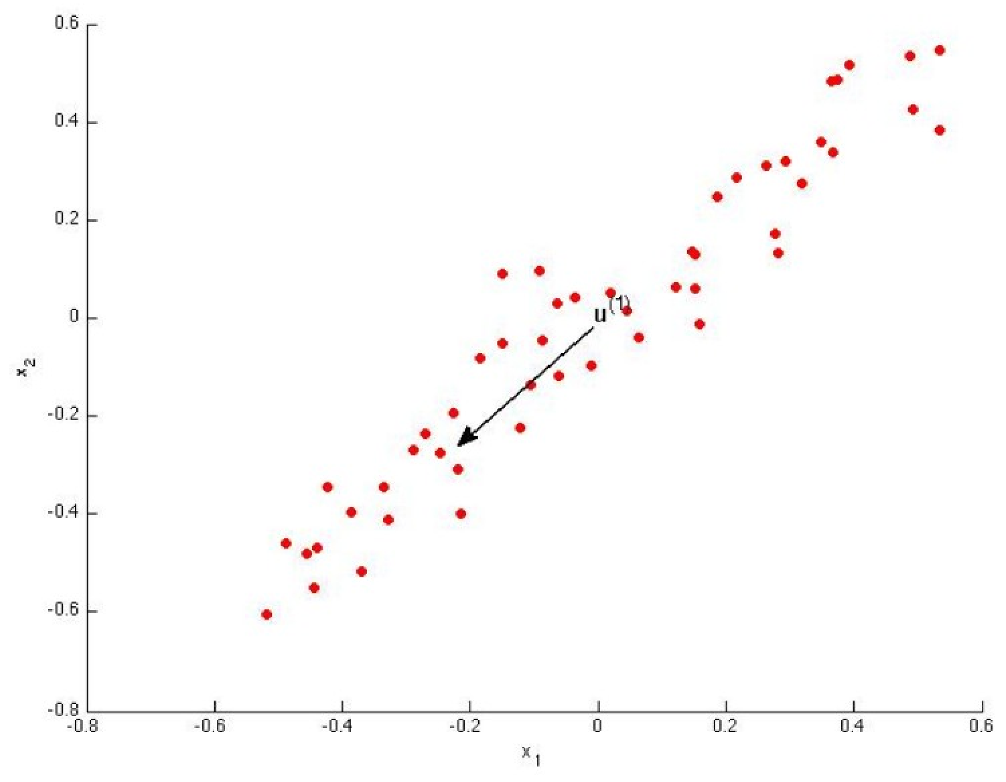
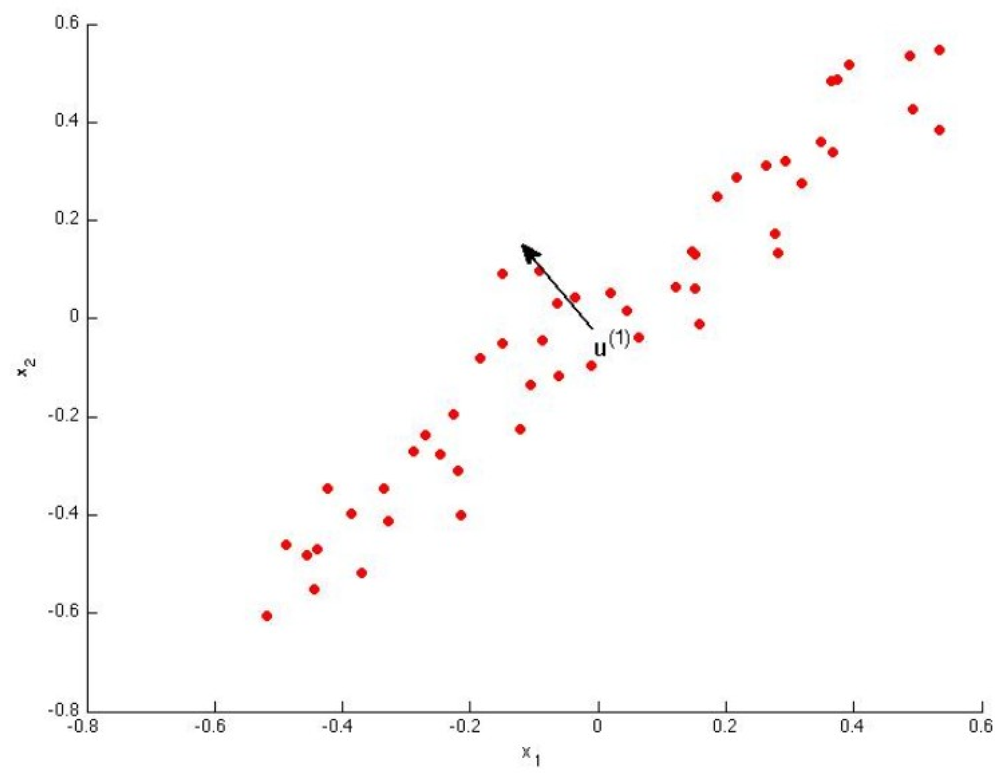
1. Consider the following 2D dataset: **1 point**

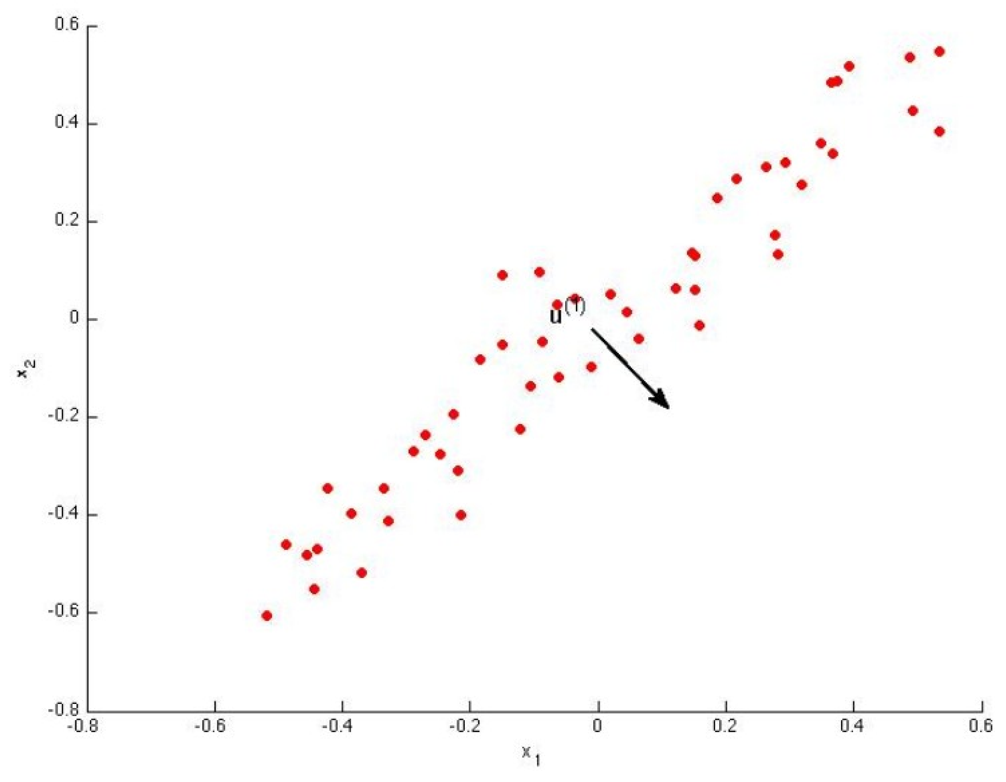


Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).

- [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.)

- ○ Use the elbow method.
- ◉ Choose k to be the smallest value so that at least 99% of the variance is retained.
- ○ Choose k to be the largest 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).

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**

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

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

- [ ] PCA is susceptible to local optima; trying multiple random initializations may help.
- [ ] Given only $z^{(i)}$ and U_{reduce}, there is no way to reconstruct any reasonable approximation to $x^{(i)}$.
- [x] Given input data $x \in \mathbb{R}^n$, it makes sense to run PCA only with values of k that satisfy $k \leq n$. (In particular, running it with $k = n$ is possible but not helpful, and $k > n$ does not make sense.)
- [x] 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.

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.
- [ ] Clustering: To automatically group examples into coherent groups.
- [ ] To get more features to feed into a learning algorithm.
- [x] Data compression: Reduce the dimension of your input data $x^{(i)}$, which will be used in a supervised learning algorithm (i.e., use PCA so that your supervised learning algorithm runs faster).