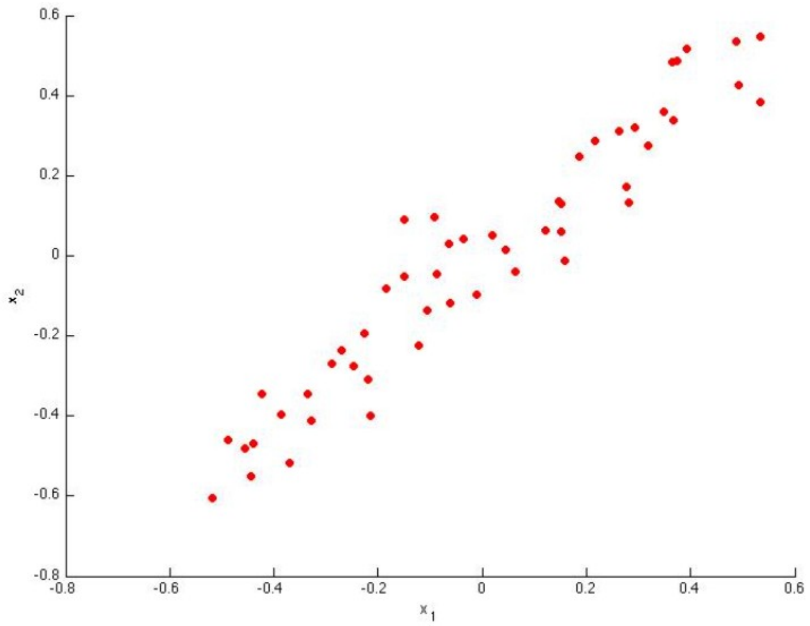
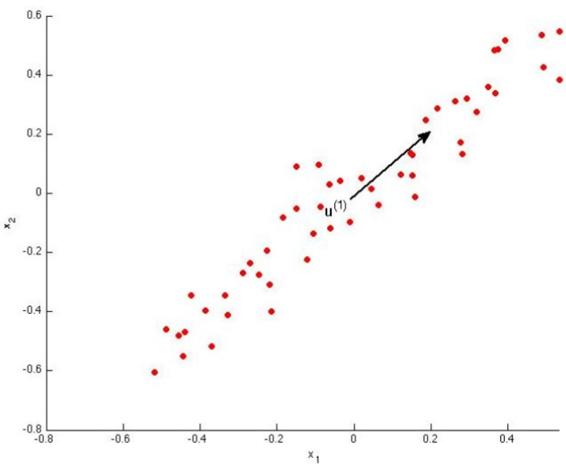
1. Consider the following 2D dataset:

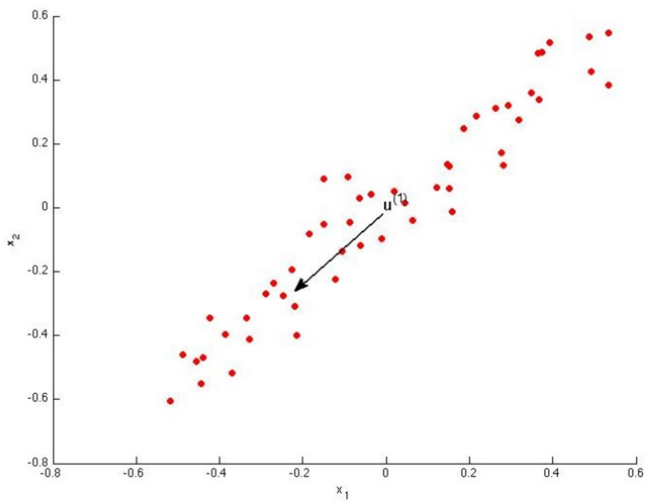
1 / 1 point



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] 

Correct
The maximal variance is along the y = x line, so this option is correct.

- [x] 

Correct
The maximal variance is along the y = x line, so the negative vector along that line is correct for the first principal component.

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

1 / 1 point

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

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

Correct
This is correct, as it maintains the structure of the data while maximally reducing its dimension.

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 / 1 point

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

Correct
This is the correct formula.

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

1 / 1 point

- [ ] 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.

Correct
Feature scaling prevents one feature dimension from becoming a strong principal component only because of the large magnitude of the feature values (as opposed to large variance on that dimension).

- [ ] PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).
- [x] Given an input $x \in \mathbb{R}^n$, PCA compresses it to a lower-dimensional vector $z \in \mathbb{R}^k$.

Correct
PCA compresses it to a lower dimensional vector by projecting it onto the learned principal components.

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

1 / 1 point

- [ ] Clustering: To automatically group examples into coherent groups.
- [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).

Correct
If your learning algorithm is too slow because the input dimension is too high, then using PCA to speed it up is a reasonable choice.

- [x] Data visualization: Reduce data to 2D (or 3D) so that it can be plotted.

Correct
This is a good use of PCA, as it can give you intuition about your data that would otherwise be impossible to see.

- [ ] To get more features to feed into a learning algorithm.