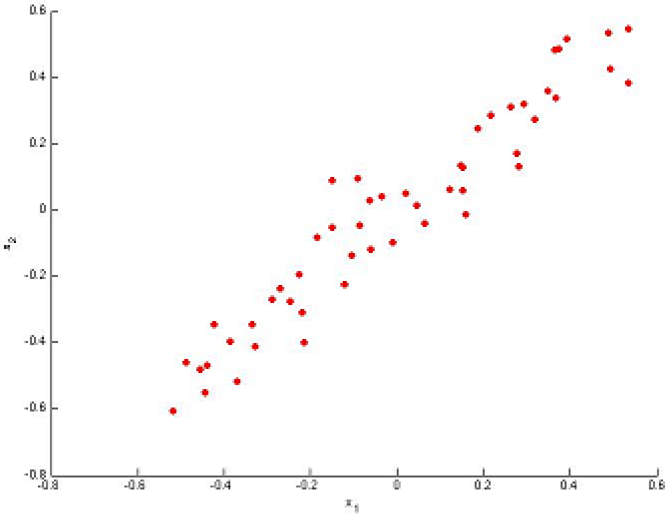
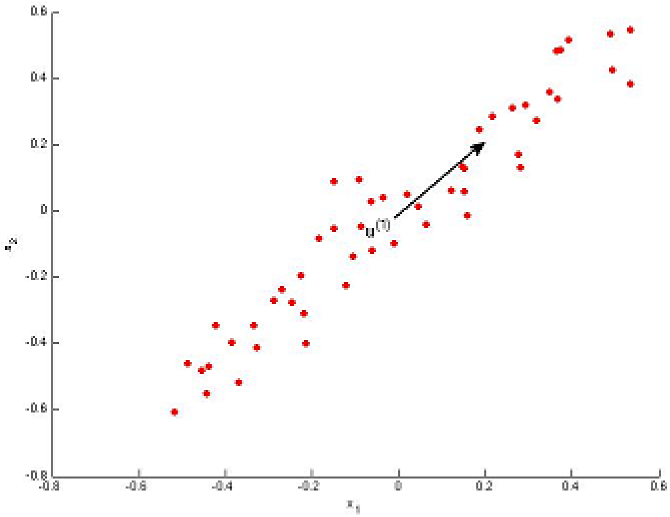
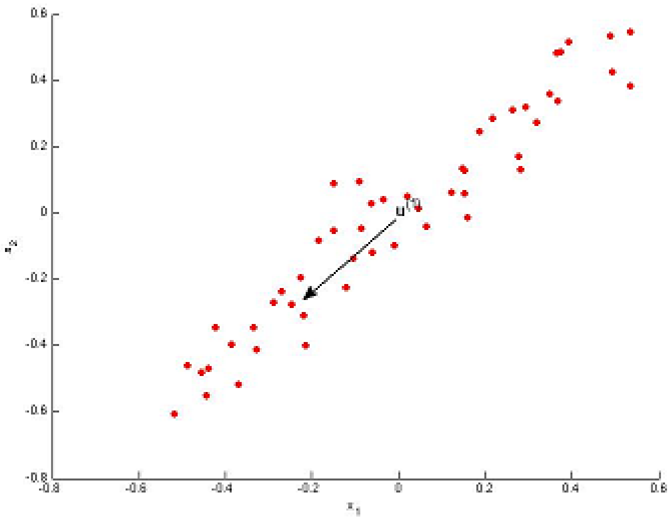
Principal Component Analysis

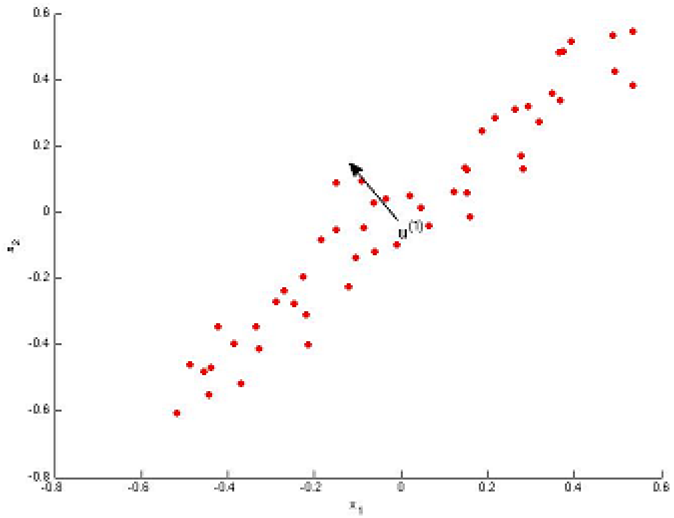
1. Consider the following 2D dataset:

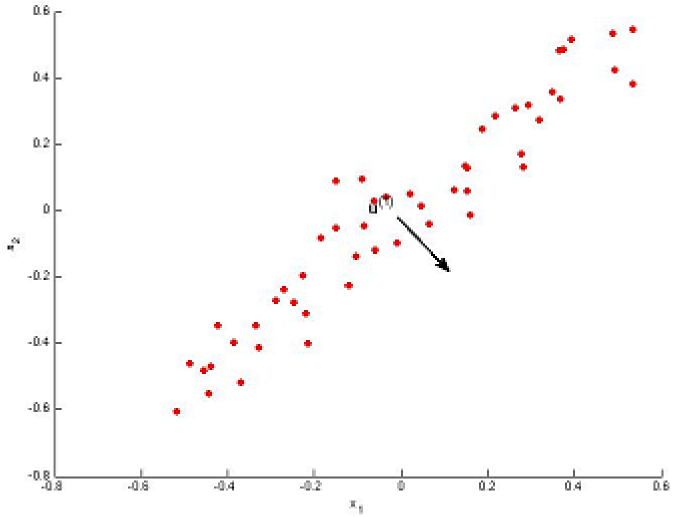


Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).





- 

- 

Option 1 - Explanation: The maximal variance is along the y = x line, so this option is correct.

Option 2 - Explanation: 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? (Recall that n is the dimensionality of the input data and m is the number of input examples).

- Choose the value of k that minimizes the approximation error $\frac{1}{m}\sum_{i=1}^{m}||x^{(i)} - x_{\text{approx}}^{(i)}||^2$
- Choose k to be 99% of n (i.e., k = 0.99 $*$ n, rounded to the nearest integer).
- Choose k to be the smallest value so that at least 1% of the variance is retained.

> Choose k to be the smallest value so that at least 99% of the variance is retained.

Explanation: 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?

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

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

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

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

> Given an input $x \in \mathbb{R}^n$, PCA compresses it to a lower-dimensional vector $z \in \mathbb{R}^k$.

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

- Feature scaling is not useful for PCA, since the eigenvector calculation (such as using Octave's **svd(Sigma)** routine) takes care of this automatically.

> If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA.

Explanation: 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).

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

- Preventing overfitting: Reduce the number of features (in a supervised learning problem), so that there are fewer parameters to learn.

> Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space.

Explanation: If memory or disk space is limited, PCA allows you to save space in exchange for losing a little of the data's information. This can be a reasonable tradeoff.

> Data visualization: Reduce data to 2D (or 3D) so that it can be plotted.

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.