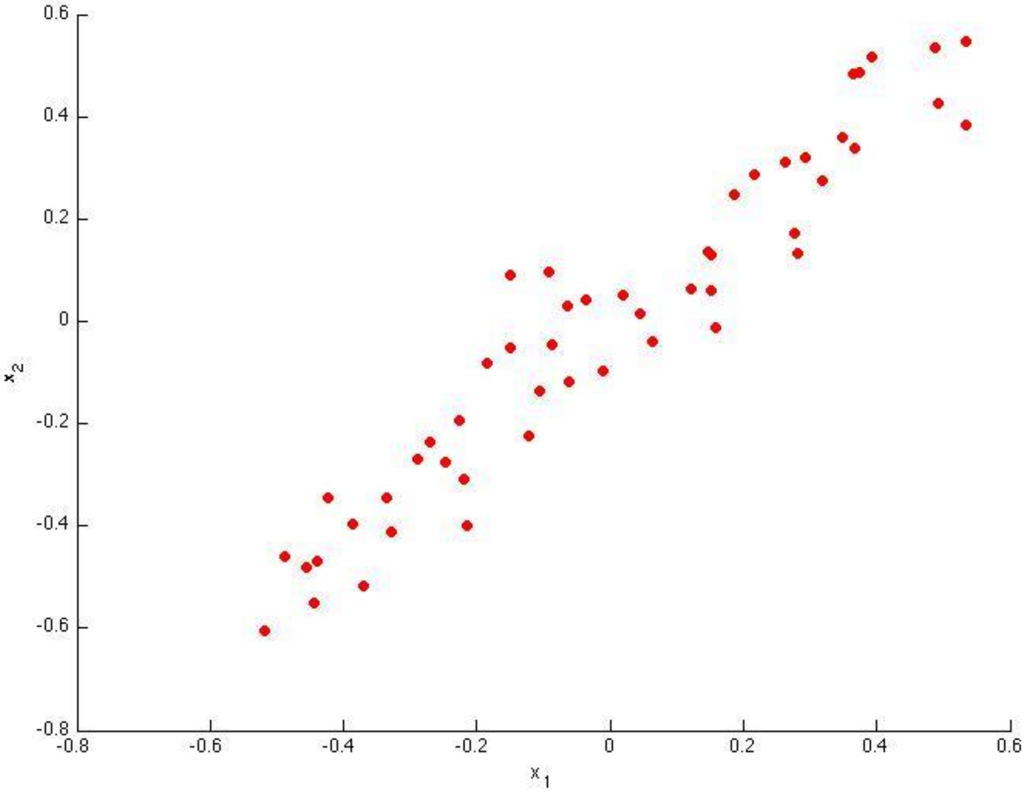
Quiz

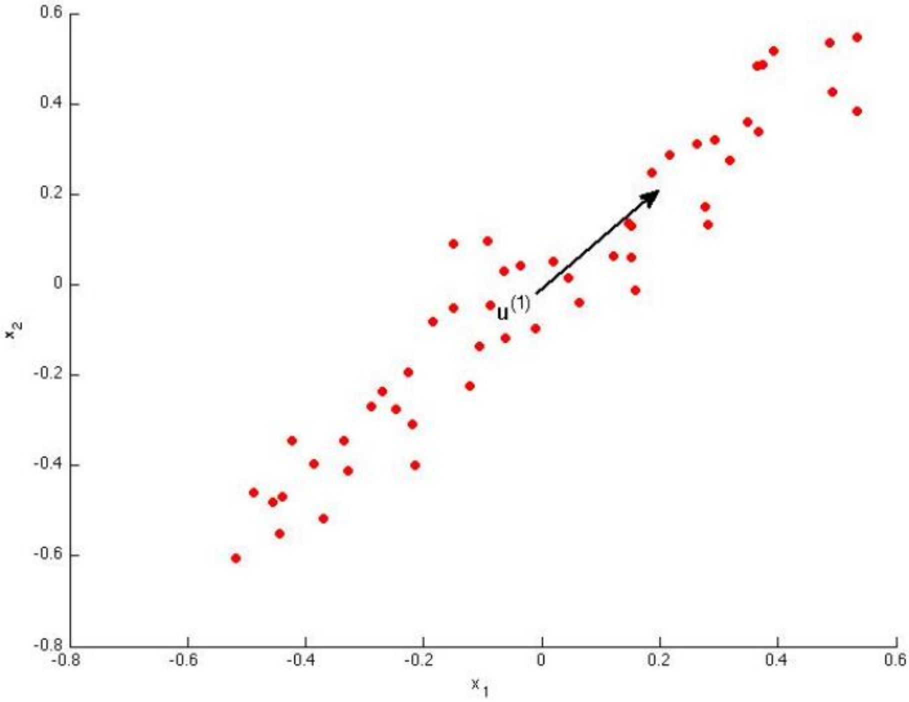
Principal Component Analysis

1. Consider the following 2D dataset:

1 point

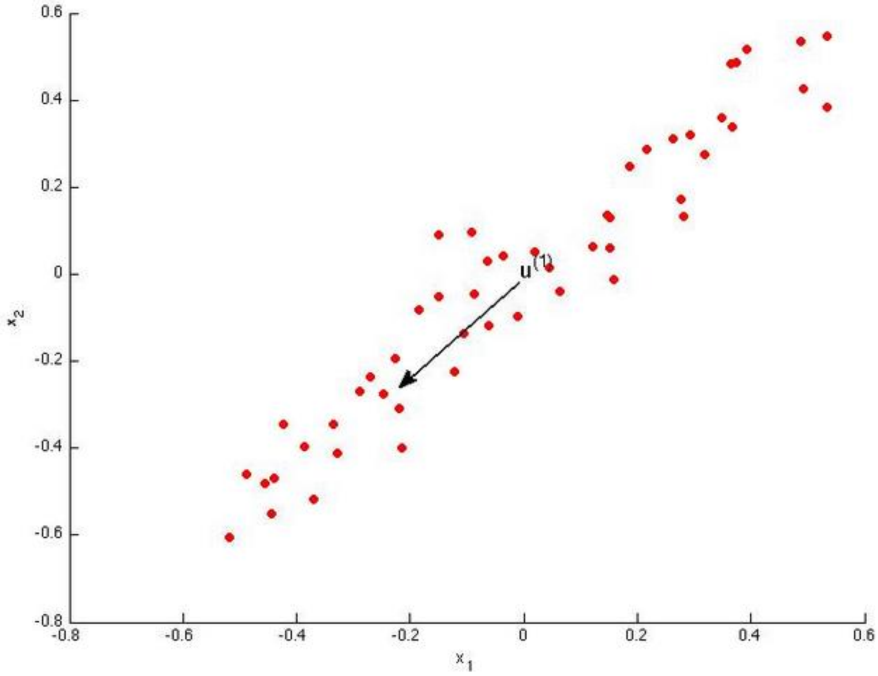


Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).



Correct

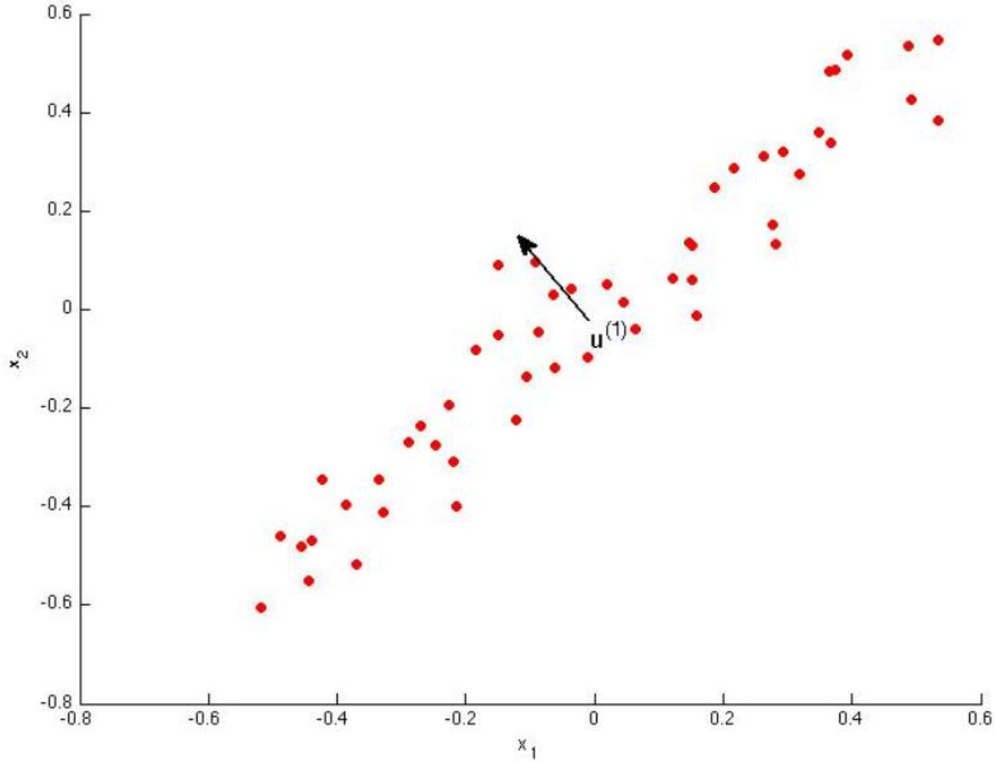
The maximal variance is along the y = x line, so this option is correct.



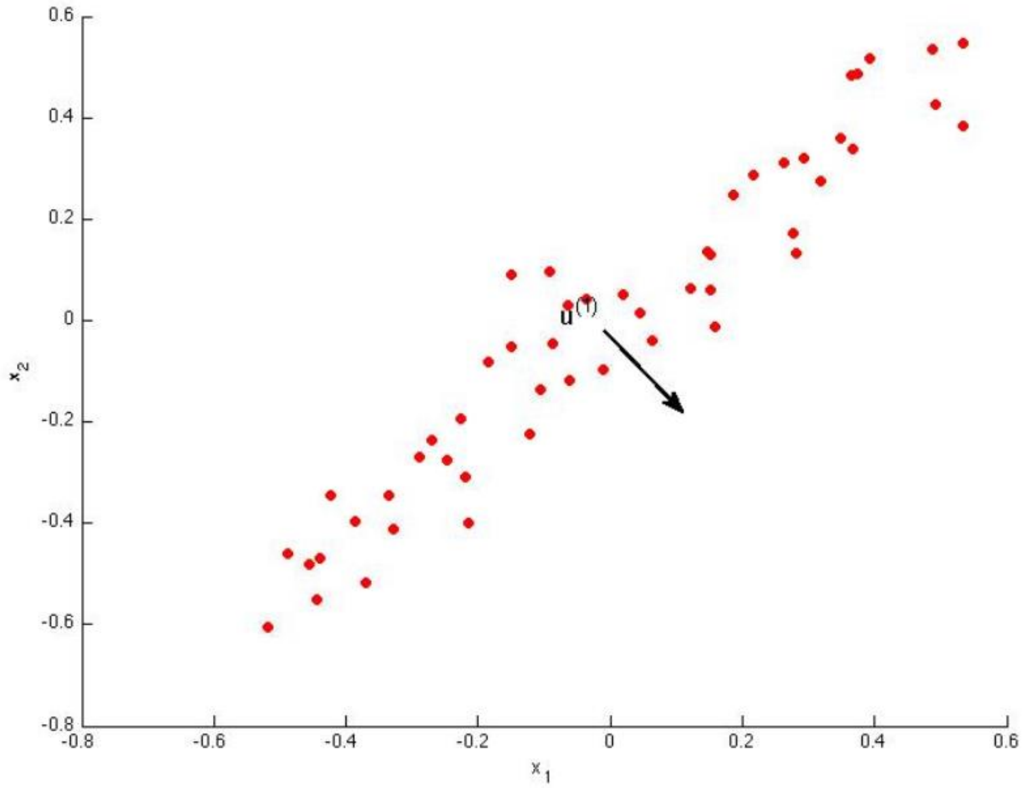
Correct

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

☐



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 / 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 largest value so that at least 99% of the variance is retained
- ◉ 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).

✓ **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

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

✓ **Correct**

This is the correct formula.

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

1 / 1 point

- [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.

✓ **Correct**

If you do not perform mean normalization, PCA will rotate the data in a possibly undesired way.

- [ ] 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.)

✓ **Correct**

The reasoning given is correct: with $k = n$, there is no compression, so PCA has no use.

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

1 / 1 point

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

- [ ] Preventing overfitting: Reduce the number of features (in a supervised learning problem), so that there are fewer parameters to learn.
- [x] Data compression: Reduce the dimension of your data, so that it takes up less memory / disk space.

Correct

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.