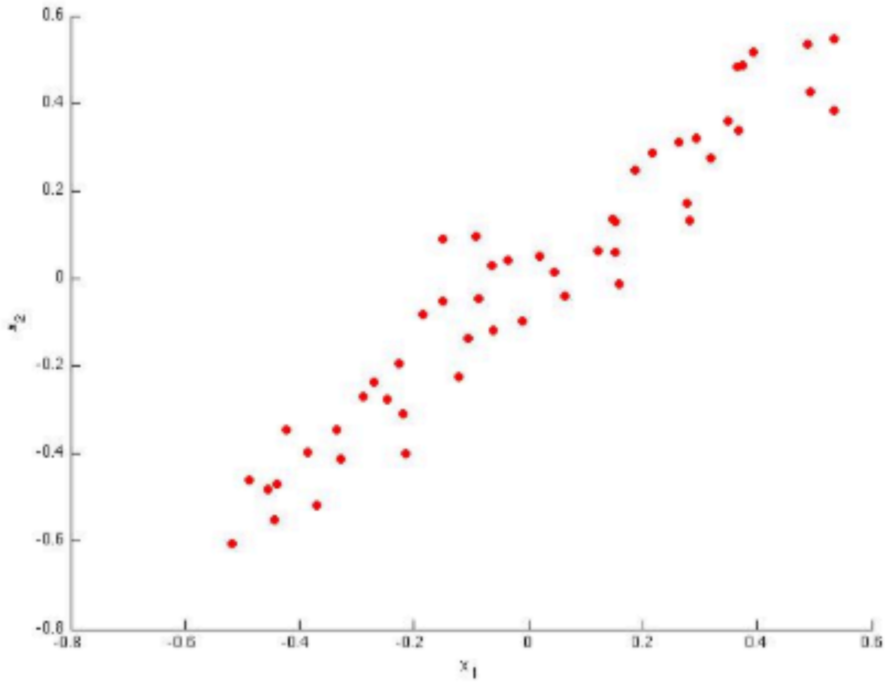
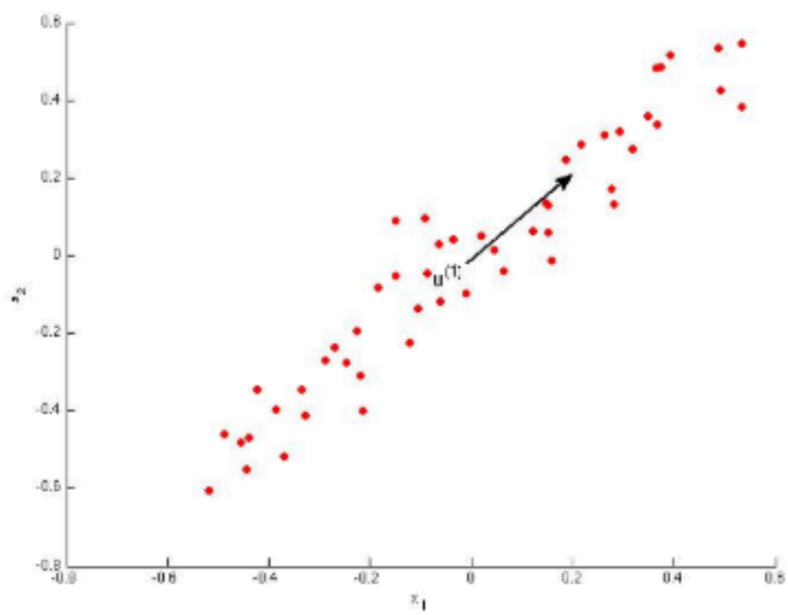
1. Consider the following 2D dataset:

1 / 1 point

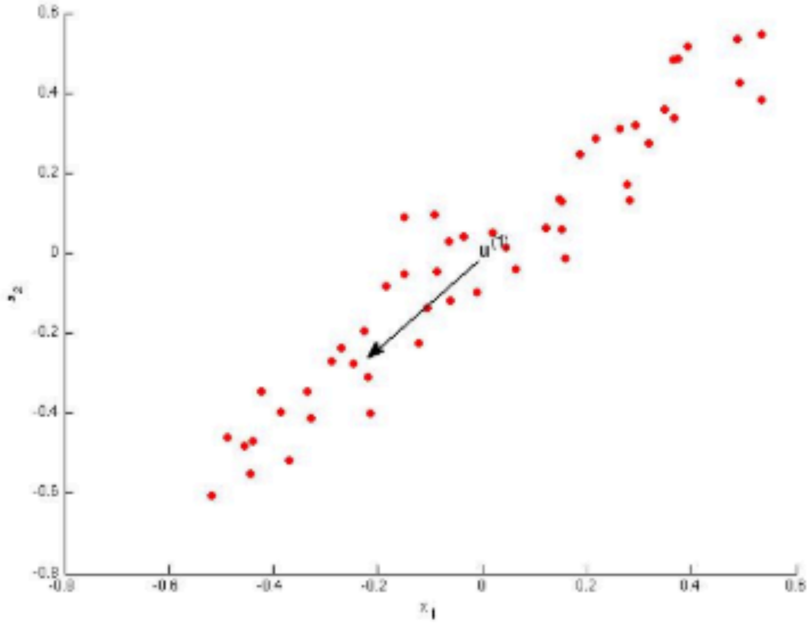


Which of the following figures correspond to possible values that PCA may return for $u^{(1)}$ (the first eigenvector / first principal component)? Check all that apply (you may have to check more than one figure).



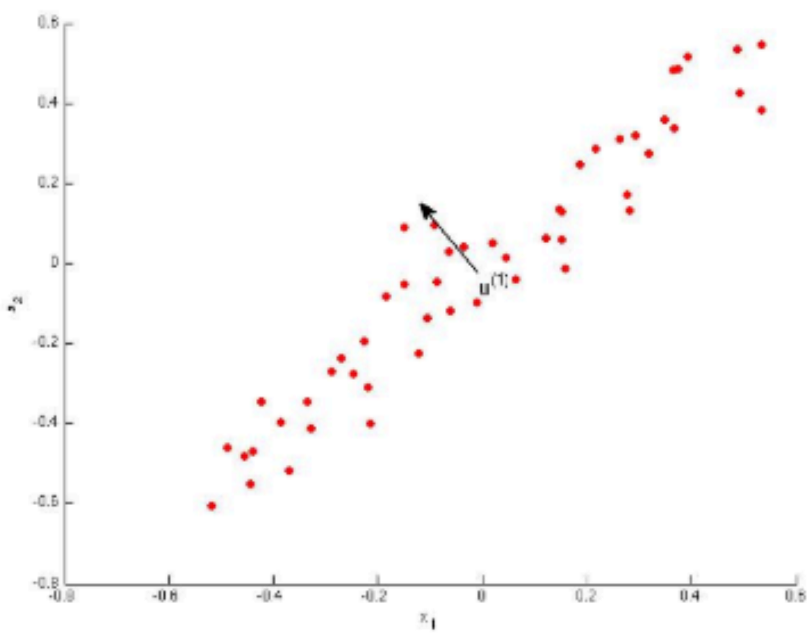
Correct

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



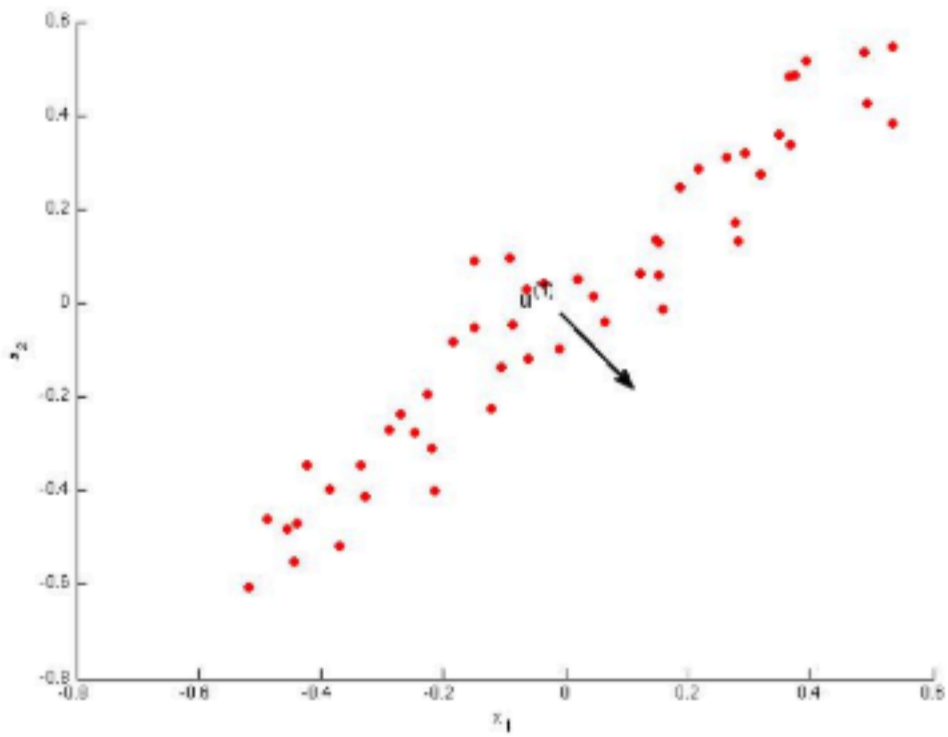
Correct

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



Un-selected is correct

☐



Un-selected is correct

0 / 1 point

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.)

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

This should not be selected

You should select the smallest such k, not the largest, as we would like to reduce the data's dimension as much as possible.

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

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

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.

Un-selected is correct

- [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.

- [ ] Given only $z^{(i)}$ and U_{reduce}, there is no way to reconstruct any reasonable approximation to $x^{(i)}$.

Un-selected is correct

---

1 / 1
point

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

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

Un-selected is correct

- [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.

- [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.

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

Un-selected is correct