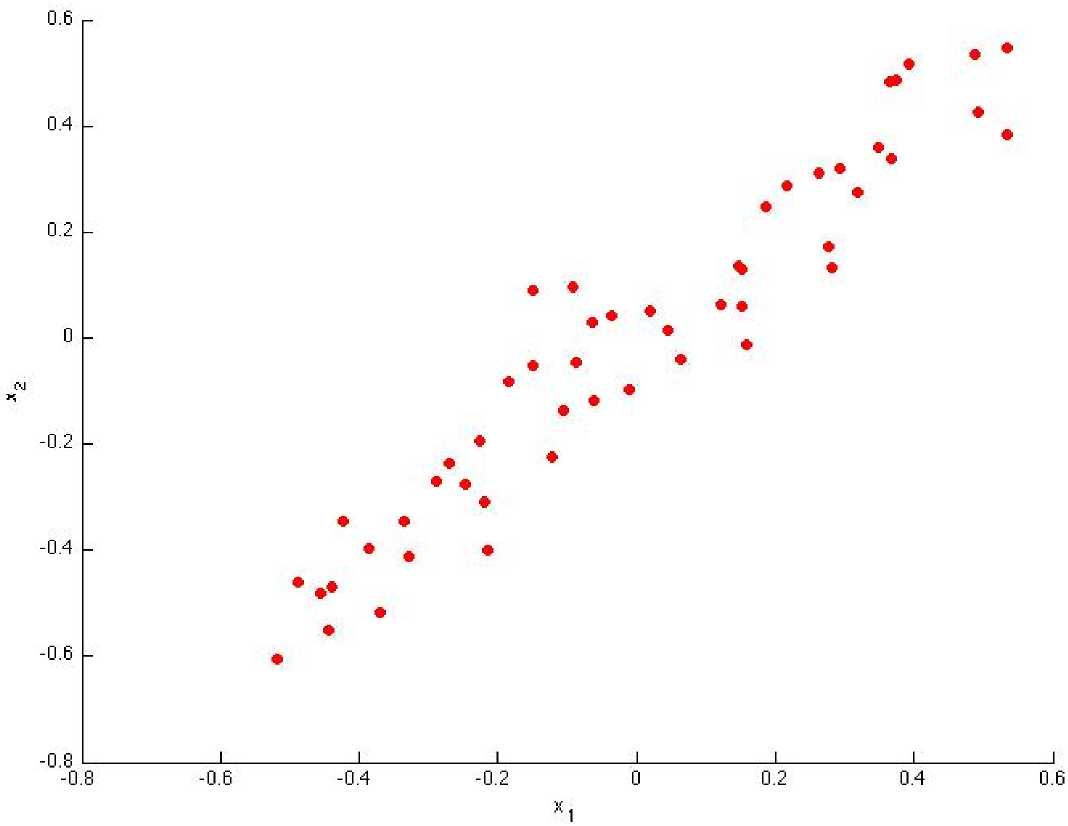
# Feedback — XIV. Principal Component Analysis

Help

You submitted this quiz on **Sun 11 May 2014 10:59 PM PDT**. You got a score of **5.00** out of **5.00**.

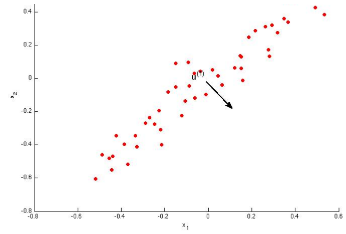
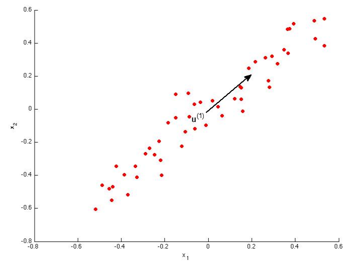
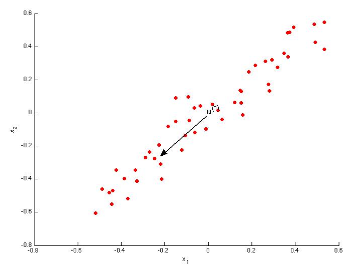
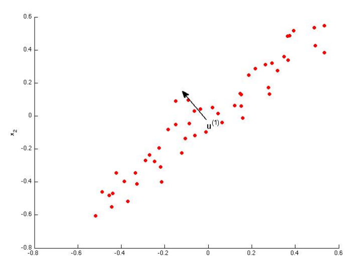
## Question 1

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).

| Your Answer | | Score | Explanation |
|---|---|---|---|
|  | ✔ | 0.25 | The first principal component is aligned with the direction of maximal variance, but this is aligned |

| | | | |
|---|---|---|---|
|  | | | with the direction of minimal variance. |
|  | ✔ | 0.25 | The maximal variance is along the y = x line, so this option is correct. |
|  | ✔ | 0.25 | The maximal variance is along the y = x line, so the negative vector along that line is correct for the first principal component. |
|  | ✔ | 0.25 | The first principal component is aligned with the direction of maximal variance, but this is aligned with the direction of minimal variance. |
| Total | | 1.00 / 1.00 | |

# Question 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.)

| Your Answer | Score | Explanation |
|---|---|---|

| | | | |
|---|---|---|---|
| ◉ Choose $k$ to be the smallest value so that at least 99% of the variance is retained. | ✔ | 1.00 | This is correct, as it maintains the structure of the data while maximally reducing its dimension. |
| ○ 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). | | | |
| Total | | 1.00 / 1.00 | |

# Question 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?

| Your Answer | | Score | Explanation |
|---|---|---|---|
| ◉ $\frac{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)} - x_{\text{approx}}^{(i)}\rvert\rvert^2}{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)}\rvert\rvert^2} \leq 0.05$ | ✔ | 1.00 | This is the correct formula. |
| ○ $\frac{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)}\rvert\rvert^2}{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)} - x_{\text{approx}}^{(i)}\rvert\rvert^2} \geq 0.05$ | | | |
| ○ $\frac{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)}\rvert\rvert^2}{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)} - x_{\text{approx}}^{(i)}\rvert\rvert^2} \leq 0.05$ | | | |
| ○ $\frac{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)} - x_{\text{approx}}^{(i)}\rvert\rvert^2}{\frac{1}{m}\sum_{i=1}^{m}\lvert\lvert x^{(i)}\rvert\rvert^2} \geq 0.95$ | | | |
| Total | | 1.00 / 1.00 | |

# Question 4

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

| Your Answer | | Score | Explanation |
|---|---|---|---|
| ☐ PCA is susceptible to local optima; trying multiple random initializations may help. | ✔ | 0.25 | PCA is a deterministic algorithm: there is no initialization and there are no local optima. |
| ☐ PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D). | ✔ | 0.25 | PCA can reduce data of dimension $n$ to any dimension $k < n$. |
| ☑ If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA. | ✔ | 0.25 | 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). |
| ☑ 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.) | ✔ | 0.25 | The reasoning given is correct: with $k = n$, there is no compression, so PCA has no use. |
| Total | | 1.00 / 1.00 | |

# Question 5

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

| Your Answer | Score | Explanation |
|---|---|---|

| | | | |
|---|---|---|---|
| ☑ 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). | ✔ | 0.25 | 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. |
| ☑ Data visualization: Reduce data to 2D (or 3D) so that it can be plotted. | ✔ | 0.25 | This is a good use of PCA, as it can give you intuition about your data that would otherwise be impossible to see. |
| ☐ As a replacement for (or alternative to) linear regression: For most learning applications, PCA and linear regression give substantially similar results. | ✔ | 0.25 | PCA is not linear regression. They have different goals (and cost functions), so they give different results. |
| ☐ To get more features to feed into a learning algorithm. | ✔ | 0.25 | PCA will reduce the number of features, not expand it. |
| Total | | 1.00 / 1.00 | |