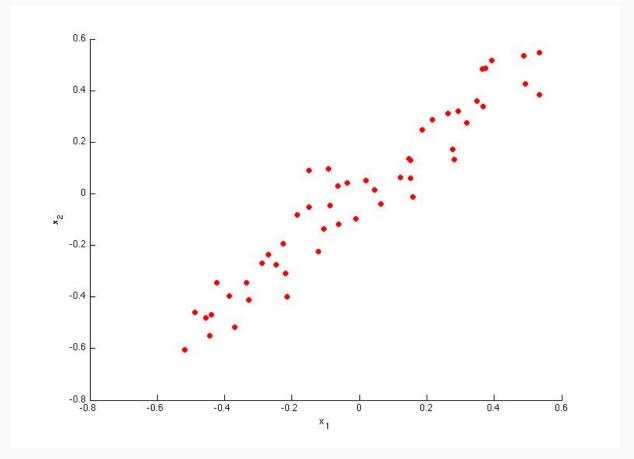
You submitted this quiz on **Sun 11 May 2014 5:51 PM IST**. You got a score of **5.00** out of **5.00**.

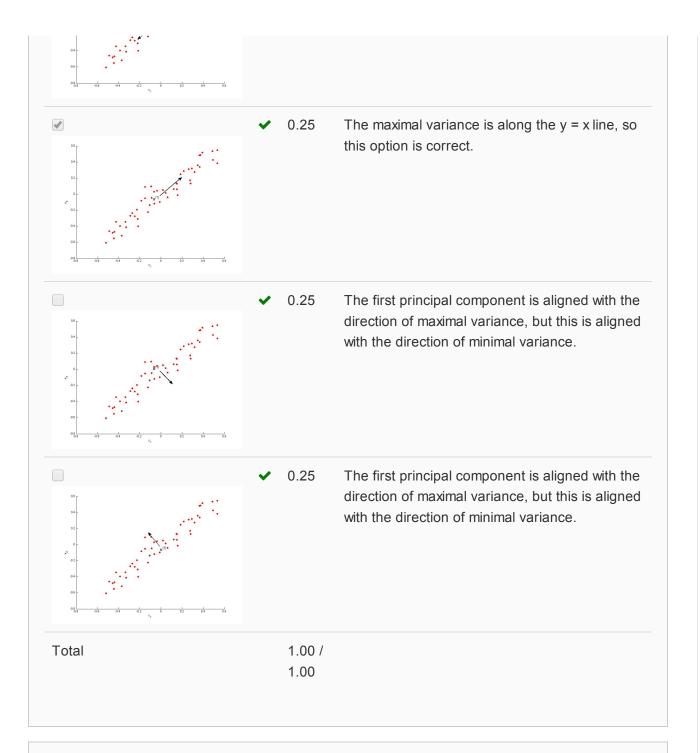


Consider the following 2D dataset:



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 maximal variance is along the y = x line, so the negative vector along that line is correct for the first principal component.
08 04 02 0 0 0 0 0 0 0 0			



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
Ouse the elbow method.	
Choose k to be 99% of m (i.e., $k=0.99*m$, rounded to the nearest integer).	
Choose the value of k that	

$rac{1}{m}\sum_{i=1}^{m}\left \left x^{(i)}-x_{ ext{approx}}^{(i)} ight ^{2}\cdot$			
ullet 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.
Гotal		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	e Expla	nation
$rac{rac{1}{m}\sum_{i=1}^{m} x^{(i)}-x_{ ext{approx}}^{(i)} ^2}{rac{1}{m}\sum_{i=1}^{m} x^{(i)} ^2}\leq 0.05$	✓ 1.00	This is	the correct formula.
$-rac{rac{1}{m}\sum_{i=1}^{m} x^{(i)} ^2}{rac{1}{m}\sum_{i=1}^{m} x^{(i)}-x_{ ext{approx}}^{(i)} ^2}\geq 0.05$			
$-rac{rac{1}{m}\sum_{i=1}^{m} x^{(i)} ^2}{rac{1}{m}\sum_{i=1}^{m} x^{(i)}-x_{ ext{approx}}^{(i)} ^2}\leq 0.95$			
$-rac{rac{1}{m}\sum_{i=1}^{m} x^{(i)} ^2}{rac{1}{m}\sum_{i=1}^{m} x^{(i)}-x_{ ext{approx}}^{(i)} ^2}\geq 0.95$			
Total	1.00 /	1.00	

Question 4

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

Your Answer		Score	Explanation
	~	0.25	PCA compresses it to a lower dimensional vector by projecting it onto the learned principal components.
■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.

✓ 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.	~	0.25	If you do not perform mean normalization, PCA will rotate the data in a possibly undesired way.
■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 $k < n$.
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 data, so that it takes up less memory / disk space.	✔ 0.25	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 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.
■To get more features to feed into a learning algorithm.	✔ 0.25	PCA will reduce the number of features, not expand it.
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.
Total	1.00 /	

1.00		