PCA Homework

Terms		
m	5	Number of instances in data set
n	2	Number of input features
p	1	Final number of principal components chosen

Original Data		
	x	*y*
p1	.2	-.3
p2	-1.1	2
p3	1	-2.2
p4	.5	-1
p5	-.6	1
mean	0	-.1

- Use PCA on the given data set to get a transformed data set with just one feature (the first principal component (PC)). Show your work along the way.
- Show what % of the total information is contained in the 1st PC.
- Do not use a PCA package to do it. You need to go through the steps yourself, or program it yourself.
- You may use a spreadsheet, Matlab, etc. to do the arithmetic for you.
- You may use any web tool or Matlab to calculate the eigenvectors from the covariance matrix.

PCA Homework

Terms		
m	5	Number of instances in data set
n	2	Number of input features
p	1	Final number of principal components chosen

Original Data		
	x	*y*
m1	.2	-.3
m2	-1.1	2
m3	1	-2.2
m4	.5	-1
m5	-.6	1
mean	0	-.1

Zero Centered Data		
	x	*y*
m1	.2	-.2
m2	-1.1	2.1
m3	1	-2.1
m4	.5	-.9
m5	-.6	1.1
mean	0	0

Covariance Matrix	
x	*y*
.715	-1.39
-1.39	2.72

EigenVectors		
x	*y*	*Eigenvalue*
.456	-.890	3.431
-.890	-.456	.0037

% total info in 1st principal component
3.431/(3.431 + . 0037) = 99.89%

Matrix A – p × n		
	x	*y*
1st PC	.456	-.890

Matrix B = Transposed zero centered Training Set - n × m					
	m1	m2	m3	m4	m5
x	.2	-1.1	1	.5	-.6
y	-.2	2.1	-2.1	-.9	1.1

$(A \times B)^T$ - *m × p New Data Set*	
	1st PC
m1	.269
m2	-2.37
m3	2.32
m4	1.03
m5	-1.25

Note that some packages might return the opposite signs on the eigenvectors. Since eigenvectors are equivalent up to a constant (e.g. -1), your final results could have opposite signs from our solution, which is fine.