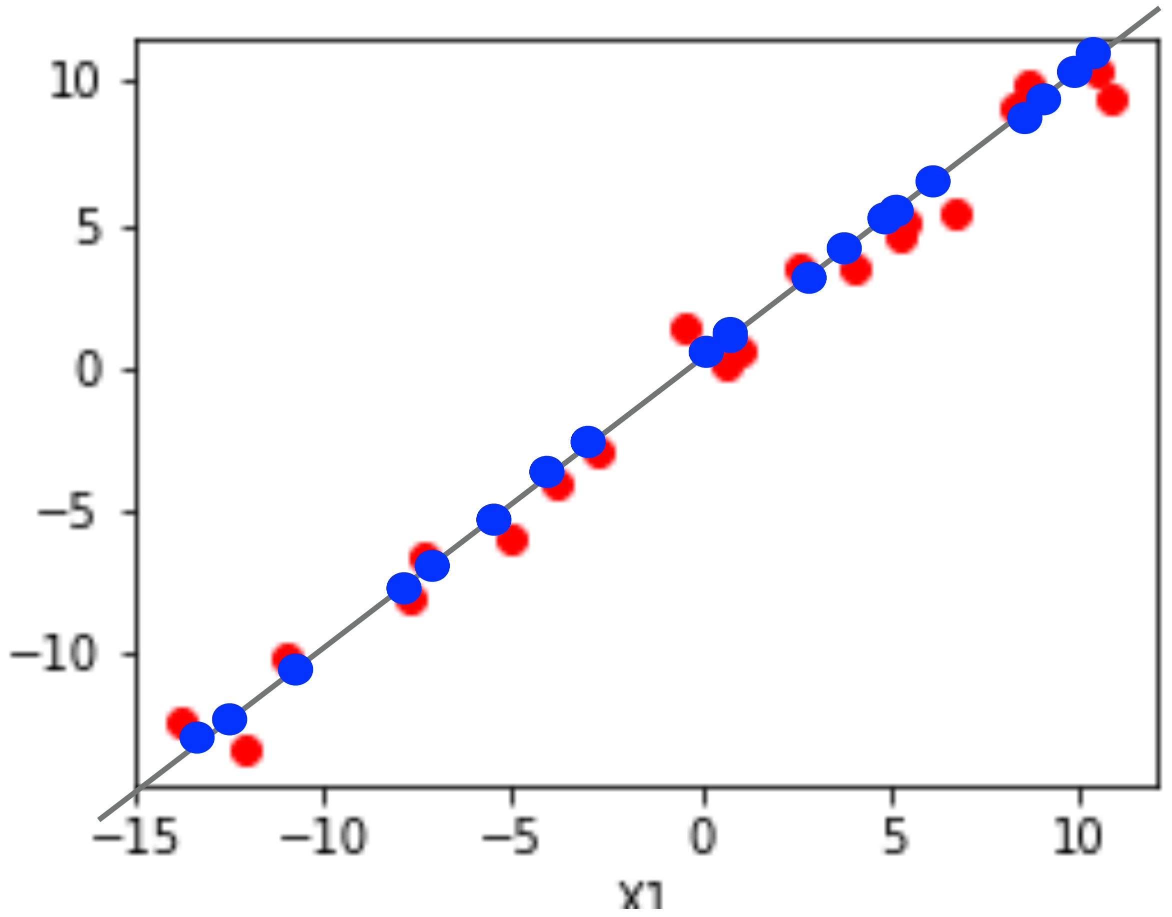
CSCI 347: Data Mining

Linear Transformations and PCA

ROTATING AXES (BASIS VECTORS) TO BETTER REPRESENT THE DATA?

- Consider the points in a data set plotted in a space where attributes represent axes:
- PCA projects the data onto a subspace that maximizes the variance preserved

$$Total\ variance = 109.45$$



aX1+bX2
18.48123029
-7.33684839
11.0687759
-12.31087461
7.69045878
5.52326352
-1.17432622
9.7911626
3.97125079
14.86526969
17.98387933
-8.60722646
-14.32291362
-0.70324989
-7.02420672
-14.71363615
-13.11970237
-.52971974
-4.22777272
-5.30481403

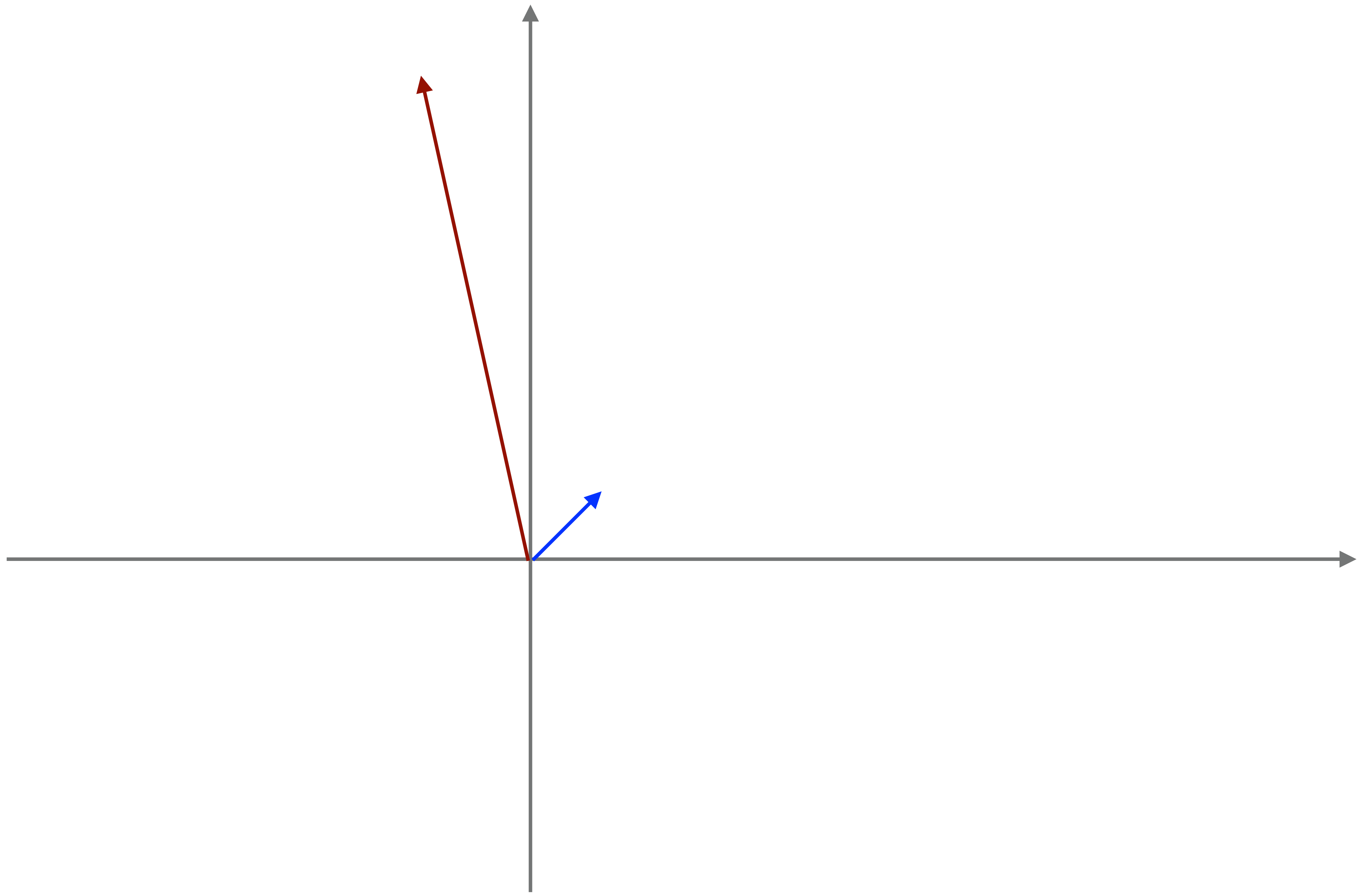
What is the best “a” and “b” = what is the best linear transformation (matrix A) to use to transform data?

$$var = 109.03$$

LINEAR TRANSFORMATIONS

Different matrices A correspond to different linear combinations:

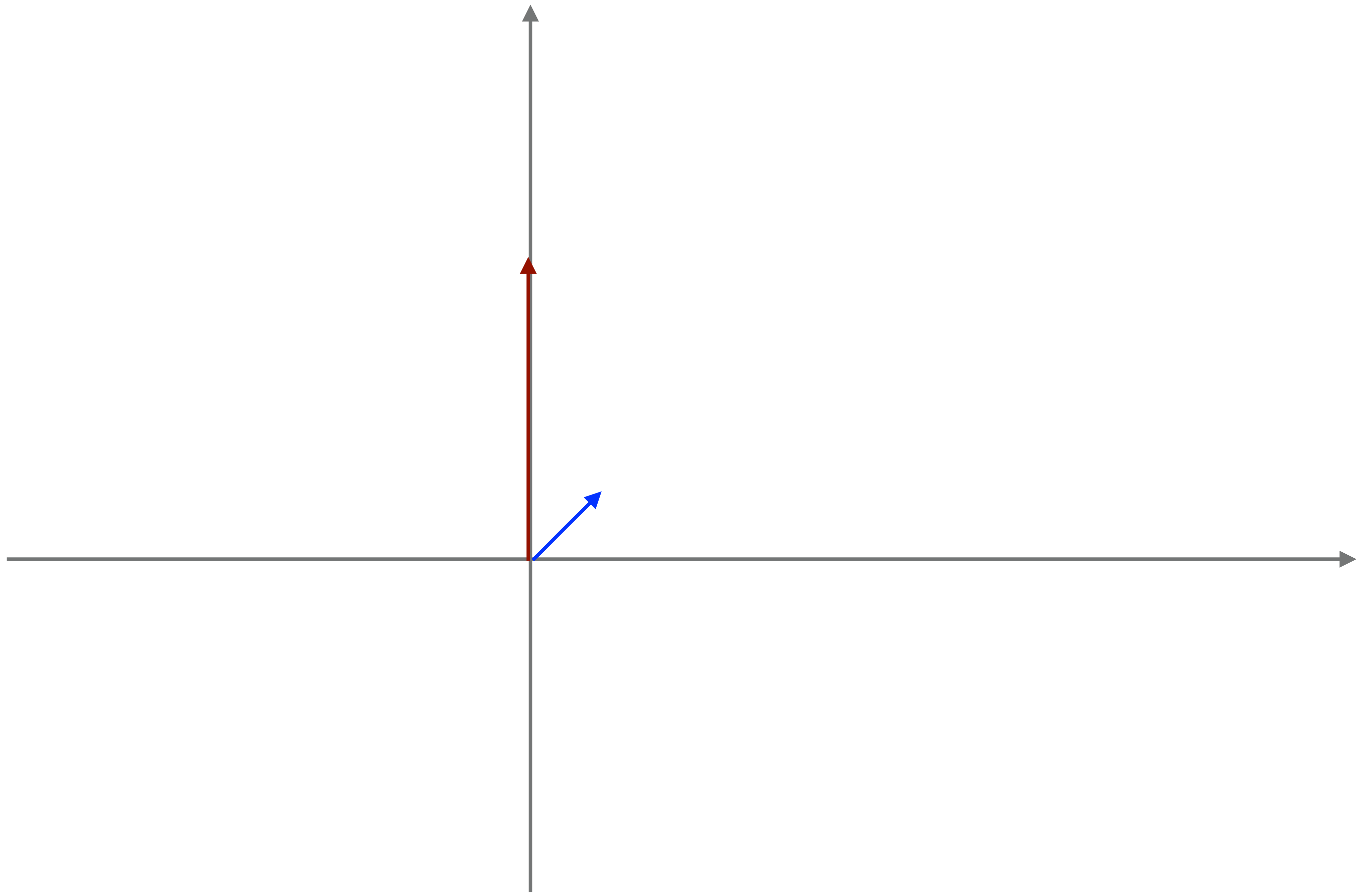
$$A = \begin{pmatrix} 1 & -2 \\ 3 & 2 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad Ax = \begin{pmatrix} 1 & -2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$$



LINEAR TRANSFORMATIONS

Different matrices A correspond to different linear combinations:

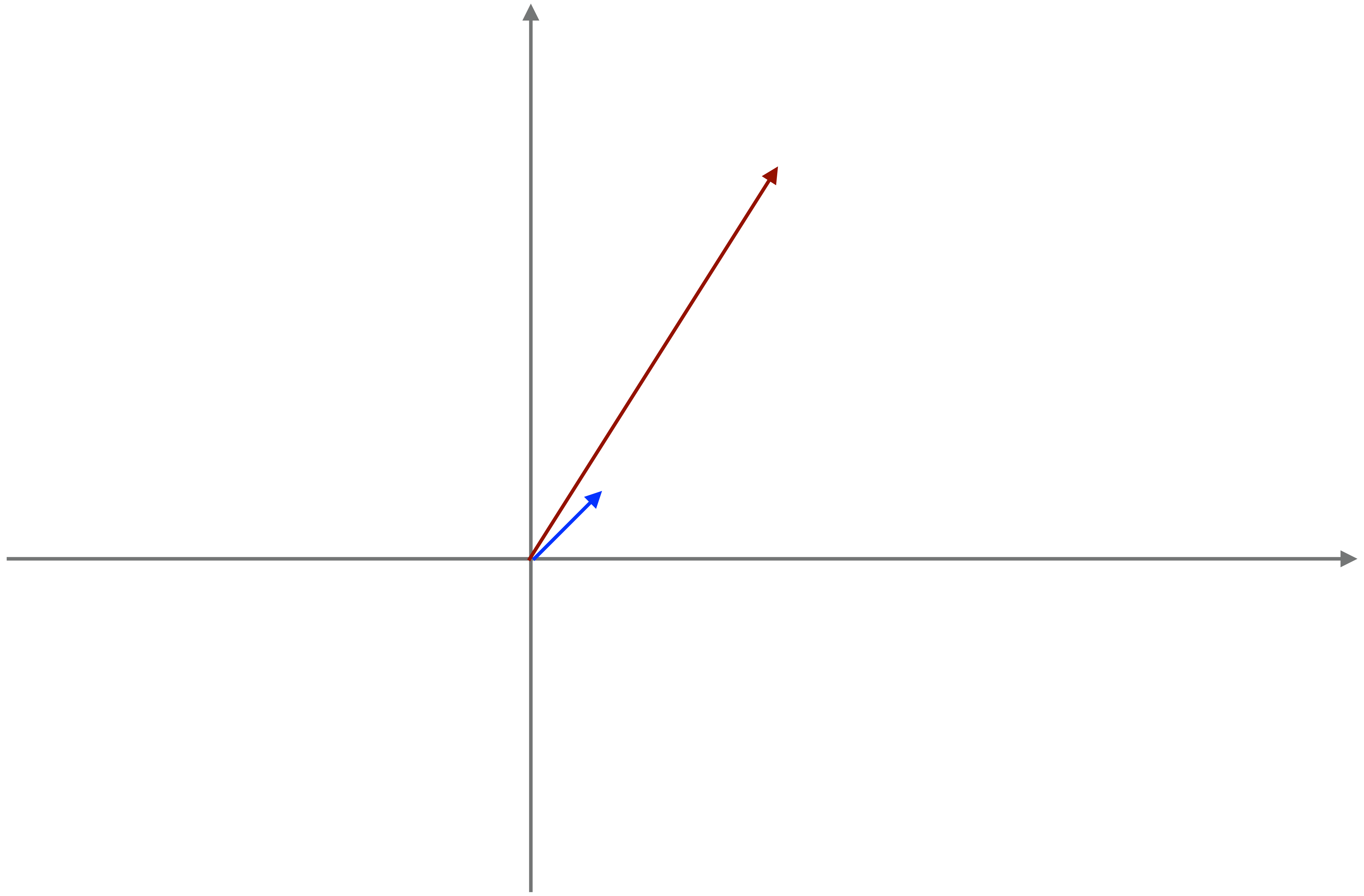
$$A = \begin{pmatrix} 2 & -2 \\ 1 & 2 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad Ax = \begin{pmatrix} 2 & -2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$



LINEAR TRANSFORMATIONS

Different matrices A correspond to different linear combinations:

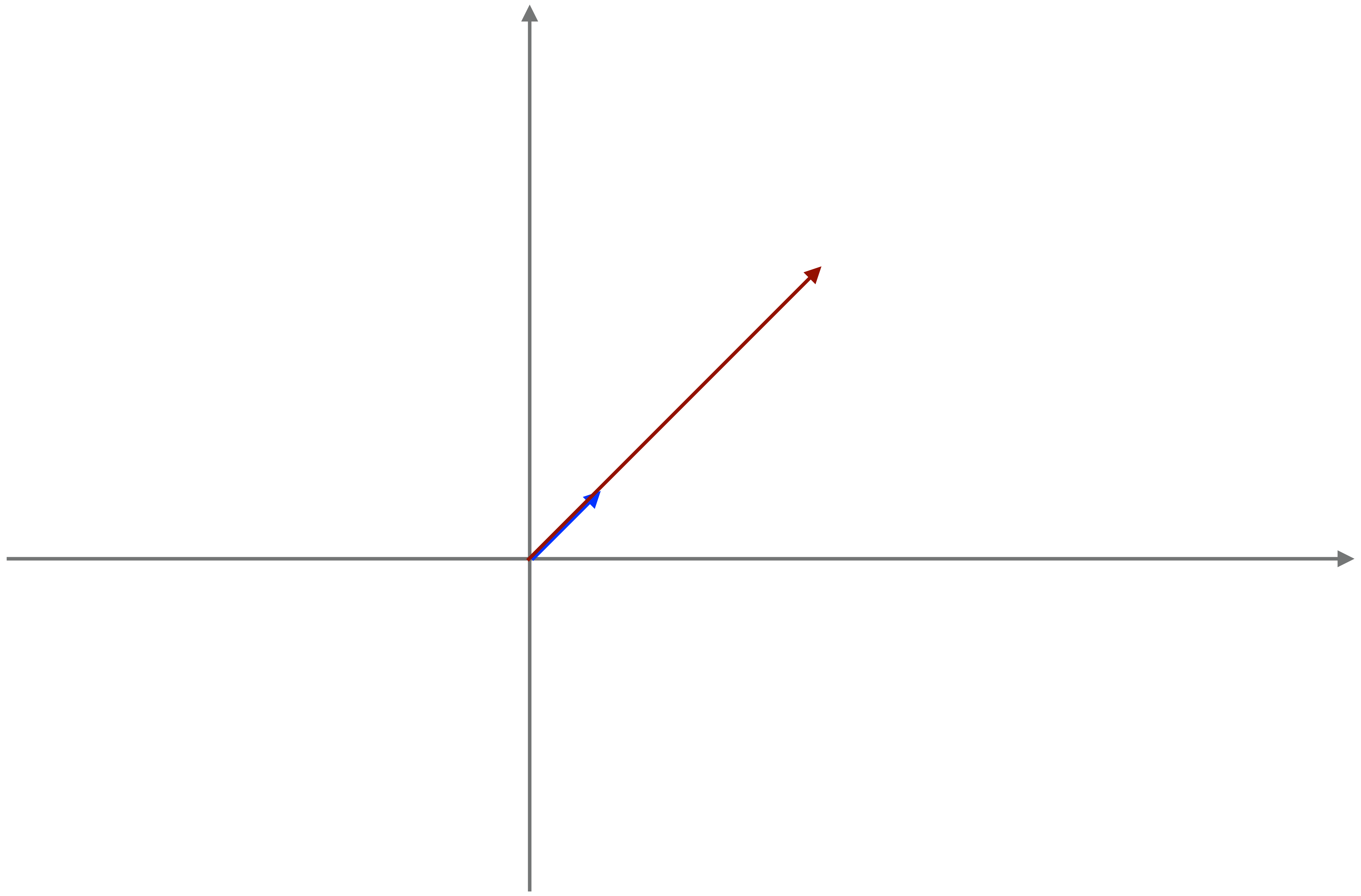
$$A = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad Ax = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$



EGIENVALUES AND EIGENVECTORS

The special vectors that, when multiplied by A, only change magnitude, are A's *eigenvectors* and the change in magnitude is the *eigenvalue*:

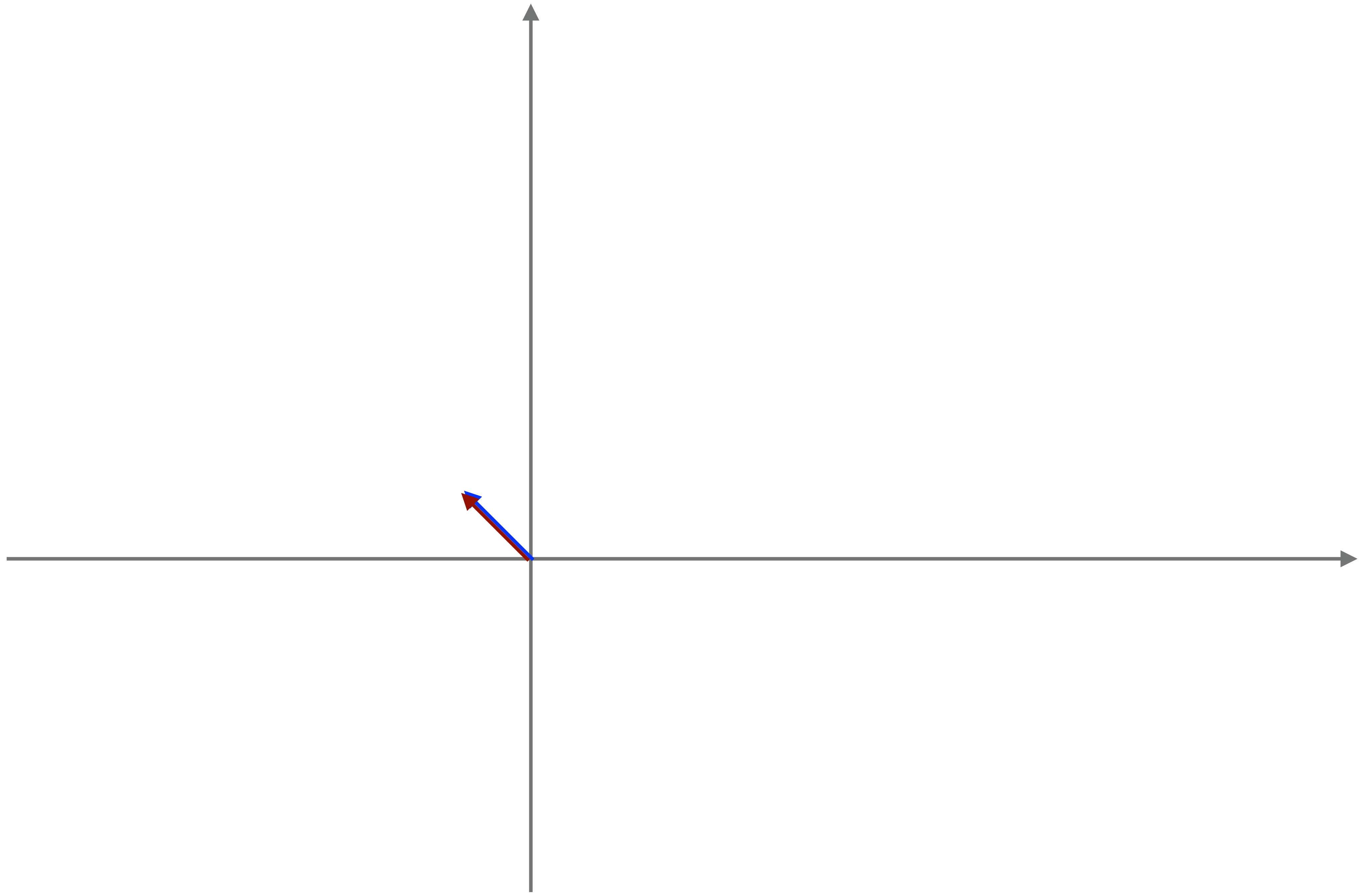
$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad Ax = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3x$$



EGIENVALUES AND EIGENVECTORS

The special vectors that, when multiplied by A, only change magnitude, are A's *eigenvectors* and the change in magnitude is the *eigenvalue*:

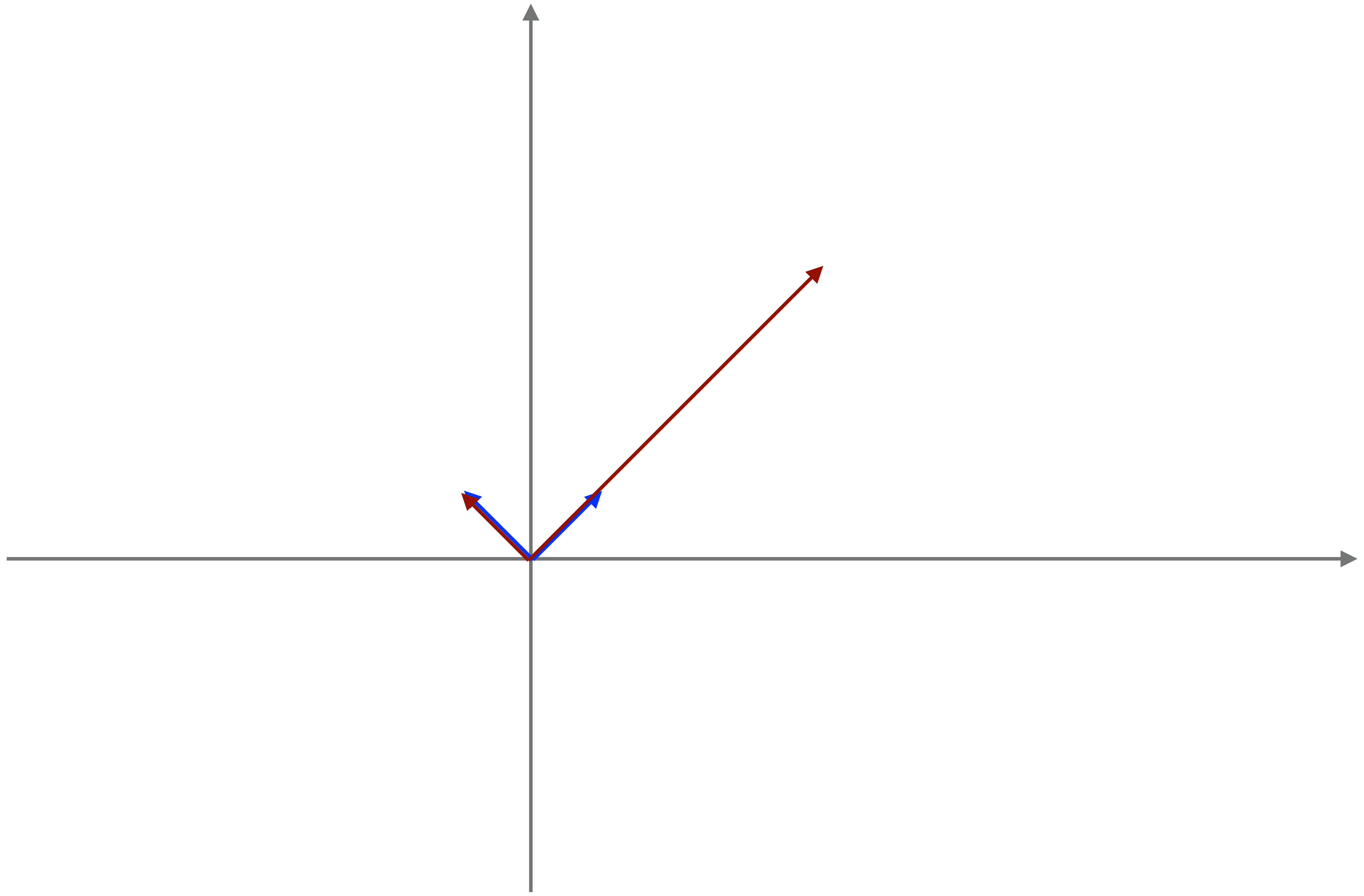
$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad x = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad Ax = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = x$$



EGIENVALUES AND EIGENVECTORS

The special vectors that, when multiplied by A, only change magnitude, are A's *eigenvectors* and the change in magnitude is the *eigenvalue*:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad e_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad Ae_1 = \lambda_1 e_1 \qquad Ae_2 = \lambda_2 e_2$$



LINEAR TRANSFORMATIONS

Consider a linear transformation of each data instance in a data matrix:

$$A = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \qquad x_1 = \begin{pmatrix} 0.2 \\ 23 \end{pmatrix} \qquad Ax = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 0.2 \\ 23 \end{pmatrix} = \begin{pmatrix} 23.4 \\ 46.4 \end{pmatrix}$$

$$D = \begin{array}{ccc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array} \quad \Rightarrow \quad \begin{array}{ccc} & 2X_1 + X_2 & 2X_1 + 2X_2 \\ x_1 & 23.4 & 46.4 \\ x_2 & 1.8 & 2.8 \\ x_3 & 4.1 & 4.6 \\ x_4 & 61.2 & 111.2 \\ x_5 & 33 & 67 \\ x_6 & 19.8 & 38.3 \\ x_7 & 13.2 & 24.2 \end{array}$$

LINEAR TRANSFORMATIONS

Consider a linear transformation of each data instance in a data matrix:

$$A = (2 \quad 1) \qquad x_1 = \begin{pmatrix} 0.2 \\ 23 \end{pmatrix} \qquad Ax = (2 \quad 1)\begin{pmatrix} 0.2 \\ 23 \end{pmatrix} = 23.4$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array} \quad \Longrightarrow \quad \begin{array}{c|c} & 2X_1 + X_2 \\ x_1 & 23.4 \\ x_2 & 1.8 \\ x_3 & 4.1 \\ x_4 & 61.2 \\ x_5 & 33 \\ x_6 & 19.8 \\ x_7 & 13.2 \end{array}$$

LINEAR TRANSFORMATIONS

Consider a linear transformation of each data instance in a data matrix:

$$A = (2 \quad 2) \qquad x_1 = \begin{pmatrix} 0.2 \\ 23 \end{pmatrix} \qquad Ax = (2 \quad 2)\begin{pmatrix} 0.2 \\ 23 \end{pmatrix} = 46.4$$

$$D = \begin{array}{ccc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array} \qquad \begin{array}{cc} & 2X_1 + 2X_2 \\ x_1 & 46.4 \\ x_2 & 2.8 \\ x_3 & 4.6 \\ x_4 & 111.2 \\ x_5 & 67 \\ x_6 & 38.8 \\ x_7 & 24.2 \end{array}$$

TRANSFORMATIONS USING EIGENVECTORS OF Σ

- *Start with a mean-centered data set Z*
- *Find its covariance matrix*
- *Find the eigenvectors and eigenvalues of that matrix*
- *Multiply each data instance by the matrix of eigenvectors to generate a new matrix of* ***linearly transformed*** *data instances*
- *The covariance matrix of this transformed data set shows that there is* ***no covariance*** *between the attributes!*

WITHOUT FURTHER ADO: PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\Sigma = \begin{matrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{matrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

$$\mu = (1.29 \quad 19.79)$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $Z = D - 1 \cdot \mu^T$

$$Z = \begin{array}{c|cc} & Z_1 & Z_2 \\ z_1 & -1.1 & 3.2 \\ z_2 & -0.9 & -18.8 \\ z_3 & 0.5 & -19.3 \\ z_4 & 4.3 & 30.2 \\ z_5 & -1.8 & 14.2 \\ z_6 & -0.9 & -0.8 \\ z_7 & -0.2 & -8.8 \end{array}$$

$$\mu_Z = (0 \quad 0)$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $Z = D - 1 \cdot \mu^T$

3. $\Sigma = \frac{1}{n-1} Z^T Z$

$$Z = \begin{array}{c|cc} & Z_1 & Z_2 \\ z_1 & -1.1 & 3.2 \\ z_2 & -0.9 & -18.8 \\ z_3 & 0.5 & -19.3 \\ z_4 & 4.3 & 30.2 \\ z_5 & -1.8 & 14.2 \\ z_6 & -0.9 & -0.8 \\ z_7 & -0.2 & -8.8 \end{array}$$

$$\mu_Z = (0 \quad 0)$$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $Z = D - 1 \cdot \mu^T$

3. $\Sigma = \frac{1}{n-1} Z^T Z$

Note that the total variance here is:

$4.14 + 321.32 = 325.46$

$$Z = \begin{array}{c|cc} & Z_1 & Z_2 \\ z_1 & -1.1 & 3.2 \\ z_2 & -0.9 & -18.8 \\ z_3 & 0.5 & -19.3 \\ z_4 & 4.3 & 30.2 \\ z_5 & -1.8 & 14.2 \\ z_6 & -0.9 & -0.8 \\ z_7 & -0.2 & -8.8 \end{array}$$

$$\mu_Z = (0 \quad 0)$$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $Z = D - 1 \cdot \mu^T$

3. $\Sigma = \frac{1}{n-1} Z^T Z$

4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$

5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

$$Z = \begin{array}{ccc} & Z_1 & Z_2 \\ z_1 & -1.1 & 3.2 \\ z_2 & -0.9 & -18.8 \\ z_3 & 0.5 & -19.3 \\ z_4 & 4.3 & 30.2 \\ z_5 & -1.8 & 14.2 \\ z_6 & -0.9 & -0.8 \\ z_7 & -0.2 & -8.8 \end{array}$$

$$\mu_Z = (0 \quad 0)$$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$(\lambda_1 \quad \lambda_2) = (322.39 \quad 3.08)$$

$$(u_1 \quad u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} = (322.39 \quad 3.08)$$

$$(u_1 \quad u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

This is the fraction of total variance that is preserved in the first r principal components!

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$
6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} = (322.39 \quad 3.08)$$

$$(u_1 \quad u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

$$f(2) = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39 + 3.08}{322.39 + 3.08} = 1$$

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $\left(\lambda_1, \lambda_2, \ldots, \lambda_d\right) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} = (322.39 \quad 3.08)$$

$$(u_1 \quad u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

$$f(2) = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39 + 3.08}{322.39 + 3.08} = 1$$

This means
over 99% of the variance
is captured in the direction of
the first principal
component!

PCA ALGORITHM

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. $Z = D - 1 \cdot \mu^T$

3. $\Sigma = \frac{1}{n-1} Z^T Z$

4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$

5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} = (322.39 \quad 3.08)$$

$(u_1$

998

So if $\alpha = 0.99$ (or anything smaller), then $r = 1$

This means over 99% of the variance is captured in the direction of the first principal component!

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

$$f(2) = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39 + 3.08}{322.39 + 3.08} = 1$$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{ccc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$
6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$(\lambda_1 \quad \lambda_2) = (322.39 \quad 3.08)$$

$$(u_1 \quad u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

$$U_1 = (u_1) = \begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$$

$$U_1^T = (u_1^T) = (-0.058 \quad -0.998)$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

$$D_{\text{transformed}} = \begin{array}{cc} & -0.058X_1 - 0.998X_2 \\ a_1 & U_1^T x_1 \\ a_2 & U_1^T x_2 \\ a_3 & U_1^T x_3 \\ a_4 & U_1^T x_4 \\ a_5 & U_1^T x_5 \\ a_6 & U_1^T x_6 \\ a_7 & U_1^T x_7 \end{array}$$

7. **Find the new coordinates a_i by linearly transforming the original data using the matrix of the first r eigenvectors of Σ:**

$U_r = (u_1 \quad \cdots \quad u_r)$, $\qquad a_i = U_r^T x_i \qquad$ for $i = 1, \ldots, n$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ \hline x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$
6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

7. Find the new coordinates a_i by linearly transforming the original data using the matrix of the first r eigenvectors of Σ:

$$U_r = (u_1 \quad \cdots \quad u_r), \qquad a_i = U_r^T x_i \qquad \text{for } i = 1, \ldots, n$$

$$a_1 = U_1^T x_1 = (-0.058 \quad -0.998)\begin{pmatrix} 0.2 \\ 23 \end{pmatrix}$$

$$U_1 = (u_1) = \begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$$

$$U_1^T = (u_1^T) = (-0.058 \quad -0.998)$$

$$D_{\text{transformed}} = \begin{array}{cc} & -0.058X_1 - 0.998X_2 \\ a_1 & U_1^T x_1 \\ a_2 & U_1^T x_2 \\ a_3 & U_1^T x_3 \\ a_4 & U_1^T x_4 \\ a_5 & U_1^T x_5 \\ a_6 & U_1^T x_6 \\ a_7 & U_1^T x_7 \end{array}$$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ \hline x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

$$a_1 = U_1^T x_1 = (-0.058 \quad -0.998) \begin{pmatrix} 0.2 \\ 23 \end{pmatrix}$$

PCA(D, α):

1. $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$
6. Choose the smallest r such that $f(r) \geq \alpha$, where:
7. Find the new coordinates a_i by linearly transforming the original data using the matrix of the first r eigenvectors of Σ:

$$(u_1 \quad u_2) = \begin{pmatrix} & 0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

$$U_1 = (u_1) = \begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$$

$$U_1^T = \left(u_1^T\right) = (-0.058 \quad -0.998)$$

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

$$D_{\text{transformed}} = \begin{array}{c|c} & -0.058X_1 - 0.998X_2 \\ \hline a_1 & -0.058(0.2) - 0.998(23) \\ a_2 & -0.058(0.4) - 0.998(1) \\ a_3 & -0.058(1.8) - 0.998(0.5) \\ a_4 & -0.058(5.6) - 0.998(50) \\ a_5 & -0.058(-0.5) - 0.998(34) \\ a_6 & -0.058(0.4) - 0.998(19) \\ a_7 & -0.058(1.1) - 0.998(11) \end{array}$$

$U_r = (u_1 \quad \cdots \quad u_r)$, $\quad a_i = U_r^T x_i \quad$ for $i = 1, \ldots, n$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{ccc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α):

$$\begin{pmatrix}\lambda_1 & \lambda_2\end{pmatrix} = (322.39 \quad 3.08)$$

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\begin{pmatrix}u_1 & u_2\end{pmatrix} = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$$

2. $Z = D - 1 \cdot \mu^T$

$$U_1 = (u_1) = \begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$$

3. $\Sigma = \frac{1}{n-1} Z^T Z$

$$U_1^T = \left(u_1^T\right) = (-0.058 \quad -0.998)$$

4. $\left(\lambda_1, \lambda_2, \ldots, \lambda_d\right) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

$$D_{\text{transformed}} = \begin{array}{cc} & -0.058X_1 - 0.998X_2 \\ a_1 & -22.97 \\ a_2 & -1.02 \\ a_3 & -0.60 \\ a_4 & -50.22 \\ a_5 & -33.90 \\ a_6 & -18.99 \\ a_7 & -11.04 \end{array}$$

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

$$f(1) = \frac{\lambda_1}{\sum_{i=1}^{2} \lambda_i} = \frac{322.39}{322.39 + 3.08} = 0.991$$

7. Find the new coordinates a_i by linearly transforming the original data using the first r principal components:

$U_r = (u_1 \quad \cdots \quad u_r)$, $\qquad a_i = U_r^T x_i \qquad$ for $i = 1, \ldots, n$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\begin{pmatrix} -0.058 & -0.998 \\ 0.998 & 0.058 \end{pmatrix}$$

2. $Z = D -$

3. $\Sigma = \frac{1}{n-1} Z$

$-0.998)$

4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) =$ eigenvalues(Σ)

5. $(u_1, u_2, \ldots, u_d) =$ eigenvectors(Σ)

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

f

322

$= 0.991$

$+ 3.08$

early transforming the original data using the

U_r

for $i = 1, \ldots, n$

Now we have a 1-dimensional representation of the data that has almost all the variance of the 2-dimensional representation!

$$D_{\text{transformed}} = \begin{array}{c|c} & -0.058X_1 - 0.998X_2 \\ a_1 & -22.97 \\ a_2 & -1.02 \\ a_3 & -0.60 \\ a_4 & -50.22 \\ a_5 & -33.90 \\ a_6 & -18.99 \\ a_7 & -11.04 \end{array}$$

$var(D_{\text{transformed}}) =$

$$\hat{\sigma}^2 = \frac{(-22.97+19.82)^2+(-1.02+19.82)^2+(-0.60+19.82)^2+(-50.22+19.82)^2+(-33.90+19.82)^2+(-18.99+19.82)^2+(-11.04+19.82)^2}{6}$$

$= 322.16$

$$\mu = \frac{-22.97 - 1.02 - 0.60 - 50.22 - 33.90 - 18.99 - 11.04}{7} = -19.82$$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α)

1\. $u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$

2\. $\begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$

3\. $\Sigma =$ … n … $(-0.058 \quad -0.998)$

4\. $(\lambda_1, \lambda_2, \ldots, \lambda_d) =$ eigenvalues(Σ)

5\. $(u_1, u_2, \ldots, u_d) =$ eigenvectors(Σ)

6\. Choose the smallest r such that $f(r) \geq \alpha$, where:

f … 322 … $= 0.991$

… $+ 3.08$

… early transforming the original data using the

U_r … or $i = 1,\ldots,n$

Remember that the total variance of the data was 325.46

$$D_{\text{transformed}} = \begin{array}{c|c} & -0.058X_1 - 0.998X_2 \\ a_1 & -22.97 \\ a_2 & -1.02 \\ a_3 & -0.60 \\ a_4 & -50.22 \\ a_5 & -33.90 \\ a_6 & -18.99 \\ a_7 & -11.04 \end{array}$$

$$var(D_{\text{transformed}}) =$$

$$\hat{\sigma}^2 = \frac{(-22.97+19.82)^2+(-1.02+19.82)^2+(-0.60+19.82)^2+(-50.22+19.82)^2+(-33.90+19.82)^2+(-18.99+19.82)^2+(-11.04+19.82)^2}{6}$$

$$= 322.16$$

$$\mu = \frac{-22.97 - 1.02 - 0.60 - 50.22 - 33.90 - 18.99 - 11.04}{7} = -19.82$$

PCA ALGORITHM

$$\Sigma = \begin{pmatrix} 4.14 & 18.42 \\ 18.42 & 321.32 \end{pmatrix}$$

$$D = \begin{array}{c|cc} & X_1 & X_2 \\ x_1 & 0.2 & 23 \\ x_2 & 0.4 & 1 \\ x_3 & 1.8 & 0.5 \\ x_4 & 5.6 & 50 \\ x_5 & -0.5 & 34 \\ x_6 & 0.4 & 19 \\ x_7 & 1.1 & 11 \end{array}$$

PCA(D, α)

1. $u_2) = \begin{pmatrix} -0.058 & -0.998 \\ -0.998 & 0.058 \end{pmatrix}$

$\begin{pmatrix} -0.058 \\ -0.998 \end{pmatrix}$

2.

$(-0.058 \quad -0.998)$

3. $\Sigma =$

n

Remember that the total variance of the data was 325.46

4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$

5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$

6. Choose the smallest r such that $f(r) \geq \alpha$, where:

f λ 322 $+ 3.08$ $= 0.991$

322.16/325.46 = 99.0

So this 1-dimensional representation contains 99% of the variance of the two-dimensional

early transforming the original data using the

U_r or $i = 1, \ldots, n$

$$D_{\text{transformed}} = \begin{array}{c|c} & -0.058X_1 - 0.998X_2 \\ a_1 & -22.97 \\ a_2 & -1.02 \\ a_3 & -0.60 \\ a_4 & -50.22 \\ a_5 & -33.90 \\ a_6 & -18.99 \\ a_7 & -11.04 \end{array}$$

$$\mu = \frac{-22.97 - 1.02 - 0.60 - 50.22 - 33.90 - 18.99 - 11.04}{7} = -19.82$$

REMEMBER THE HIGH-LEVEL DESCRIPTION OF PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction method that uses all the attributes to create a smaller number of new attributes that can represent the data well

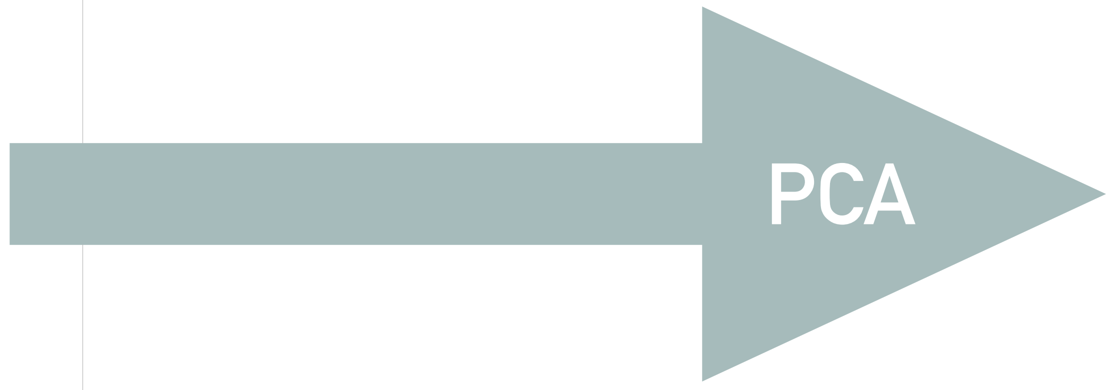
REMEMBER THE HIGH-LEVEL DESCRIPTION OF PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction method that uses all the attributes to create a smaller number of new attributes that can represent the data well

Total variance of original data = 109.45

$D =$

X1	X2
-13.75708603	-12.36844019
5.33464972	5.03907952
-7.7038562	-7.95244212
8.32381429	9.09357999
-4.99820666	-5.88569863
-3.81229004	-4.00067214
1.0380971	0.61914256
-7.34133247	-6.49878802
-2.75398505	-2.86334696
-10.97861835	-10.0367878
-12.07540131	-13.36959126
6.73139311	5.43042324
10.830517	9.4138090
-0.40654868	1.41666035
5.27546356	4.65332586
10.5040391	10.30322906
8.61616821	9.94992683
0.59588877	0.14947268
2.54124822	3.44564583
4.0360457	3.46147219



$D_{transformed} =$

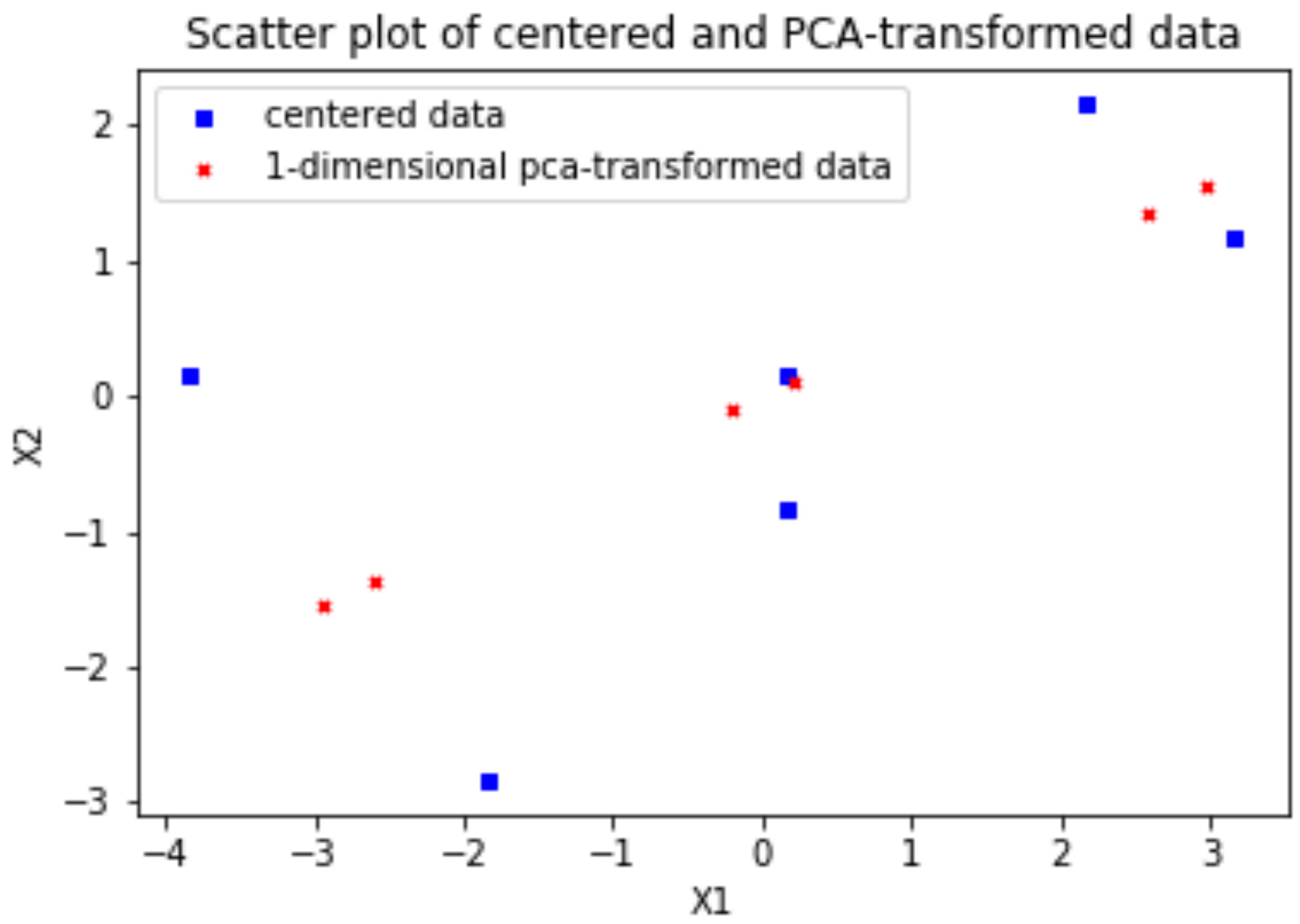
aX1+bX2
18.48123029
-7.33684839
11.0687759
-12.31087461
7.69045878
5.52326352
-1.17432622
9.7911626
3.97125079
14.86526969
17.98387933
-8.60722646
-14.32291362
-0.70324989
-7.02420672
-14.71363615
-13.11970237
-.52971974
-4.22777272
-5.30481403

Variance of transformed data = 109.03

- More specifically, new attributes are created using a **linear combination of original attributes**, where the **eigenvectors of the covariance matrix of the original data** determine the linear transformation of the original attributes. The new attributes preserve as much of the **total variance** as possible but have **no covariance between each other.**

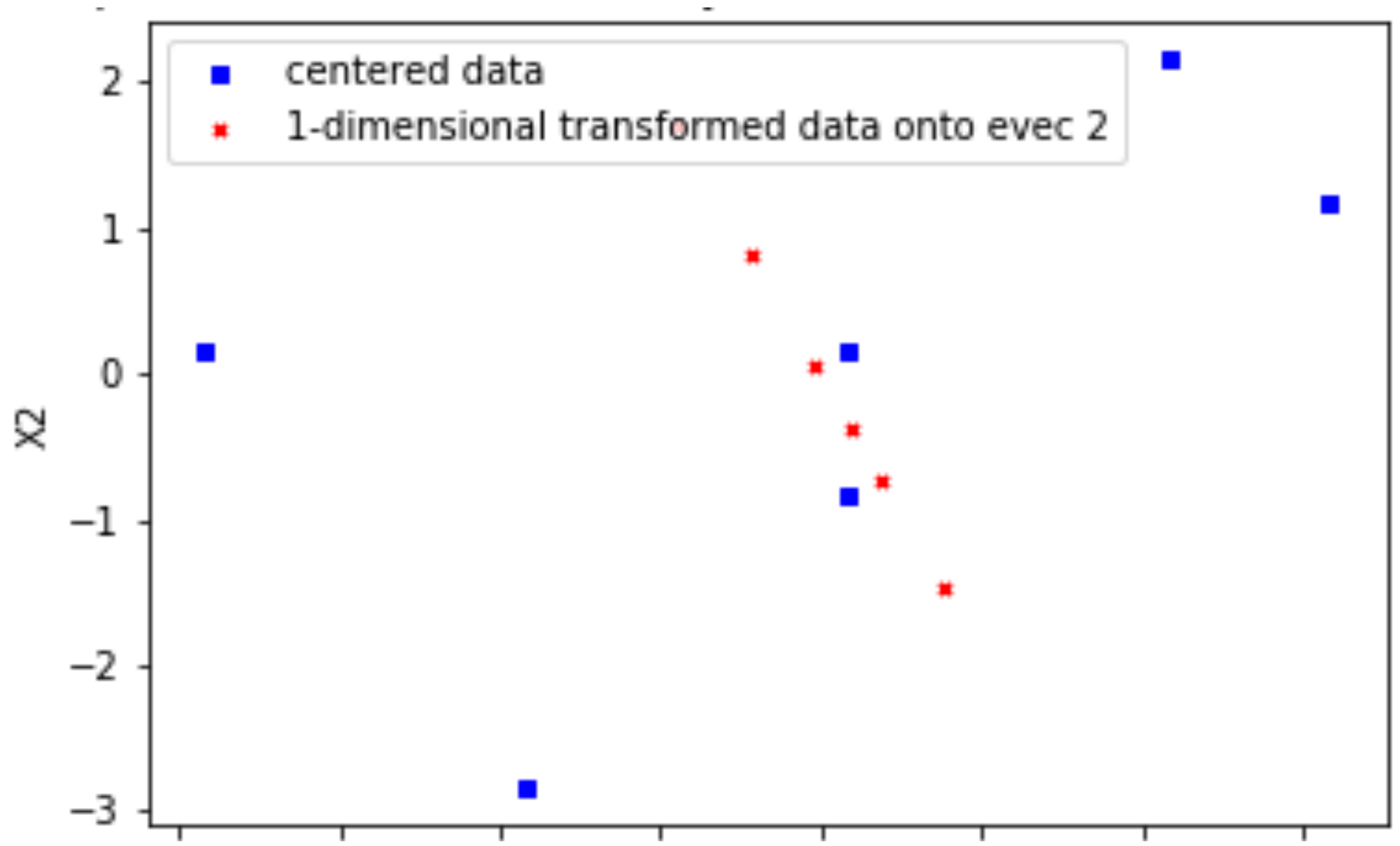
HIGH-LEVEL DESCRIPTION OF PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction method that uses all the attributes to create a smaller number of new attributes that can represent the data well
- More specifically, new attributes are created using a **linear combination of original attributes**, where the **eigenvectors of the covariance matrix of the original data** determine the linear transformation of the original attributes. The new attributes preserve as much of the **total variance** as possible but have **no covariance between each other.**
- Consider the example below:



HIGH-LEVEL DESCRIPTION OF PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction method that uses all the attributes to create a smaller number of new attributes that can represent the data well
- More specifically, new attributes are created using a **linear combination of original attributes**, where the **eigenvectors of the covariance matrix of the original data** determine the linear transformation of the original attributes. The new attributes preserve as much of the **total variance** as possible but have **no covariance between each other.**
- Consider the example below:: **what if we had used the second eigenvector?**



HIGH-LEVEL DESCRIPTION OF PRINCIPAL COMPONENT ANALYSIS

- PCA is a dimensionality reduction method that uses all the attributes to create a smaller number of new attributes that can represent the data well
- More specifically, new attributes are created using a **linear combination of original attributes**, where the **eigenvectors of the covariance matrix of the original data** determine the linear transformation of the original attributes. The new attributes preserve as much of the **total variance** as possible but have **no covariance between each other.**
- Consider the example below: **what if we had used both eigenvectors?**