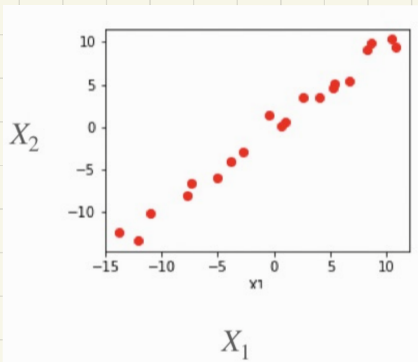
Intro to PCA

Dim Reduction As Feature Selection

If goal is to maintain total variance, how to choose?



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

Recall

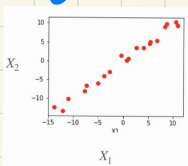
# of data dim, e.g. # cols

$$Var(D) = \sum_{d=1}^{m} \sigma_d^2$$

$$Var(D) = 109.45$$

Dim Reduction As Feature Selection

If goal is to maintain total variance, how to choose?

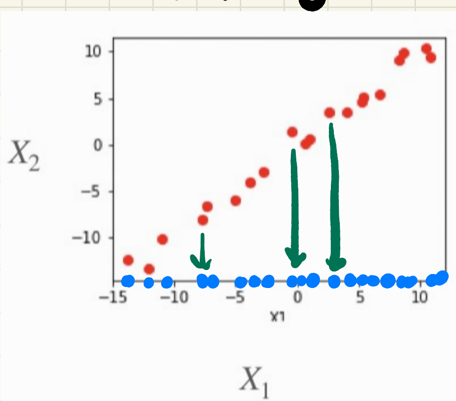


X1	X2
-13.75708603	-12.36844019
5.33464972	5.03907952
-7.7038562	-7.95244212
8.32381429	9.09357999
-4.99820666	-5.88569863
-3.81229004	-4.00067214
1.0380971	0.61914256
-7.34133247	-6.49878802
-2.75398505	-2.86334696
-10.97861835	-10.0367878
-12.07540131	-13.36959126
6.73139311	5.43042324
10.830517	9.4138090
-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

$Var(D) = 109.45$

Project Data onto space spanned by

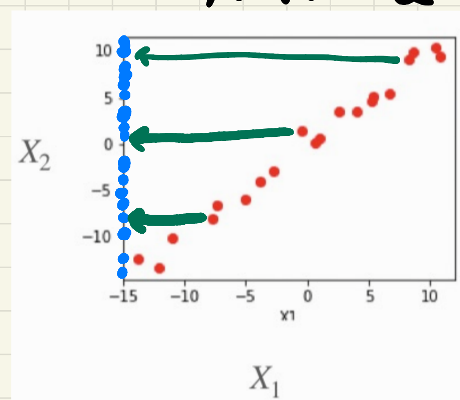
Attrib 1



X1
-13.75708603
5.33464972
-7.7038562
8.32381429
-4.99820666
-3.81229004
1.0380971
-7.34133247
-2.75398505
-10.97861835
-12.07540131
6.73139311
10.830517
-0.40654868
5.27546356
10.5040391
8.61616821
0.59588877
2.54124822
4.0360457

$\sigma_1^2 = 55.65$

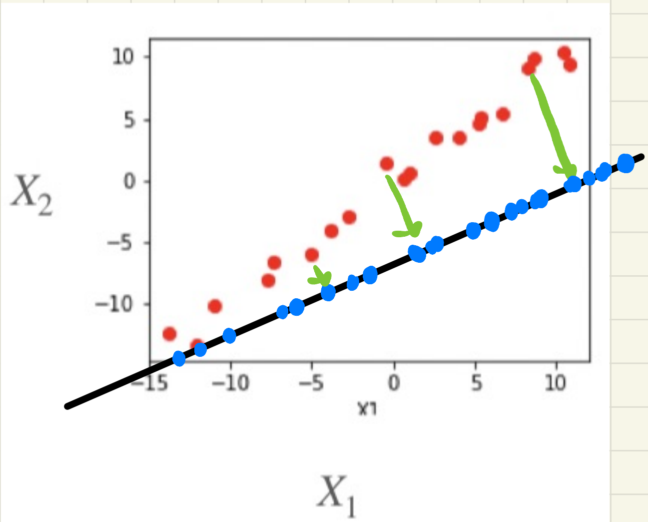
Attr. 2



X2
-12.36844019
5.03907952
-7.95244212
9.09357999
-5.88569863
-4.00067214
0.61914256
-6.49878802
-2.86334696
-10.0367878
-13.36959126
5.43042324
9.4138090
1.41666035
4.65332586
10.30322906
9.94992683
0.14947268
3.44564583
3.46147219

$\sigma_2^2 = 53.80$

Project onto some 1-d Subspace



D_1

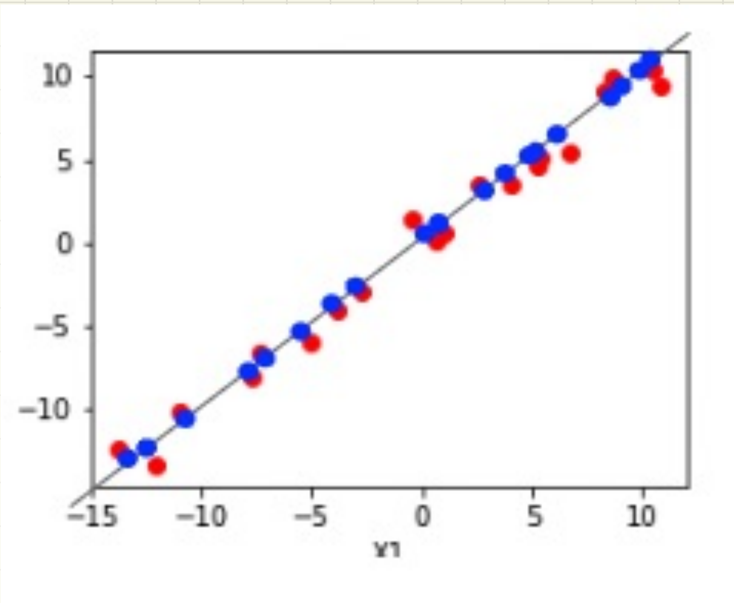
$aX_1 + bX_2$

$Var(D_1) > Var(X_1)$

$Var(D_2) > Var(X_2)$

Project onto "best" 1-d subspace

PCA projects onto subspace maximizing preserved Var



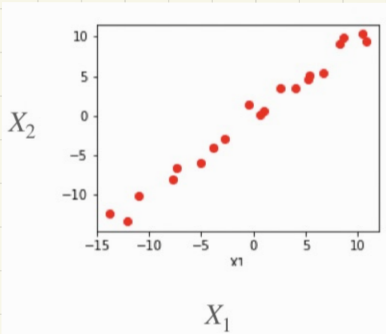
$D_{PCA} =$

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

$V(D) = 109.45$

$V(D_{PCA}) = 109.03$

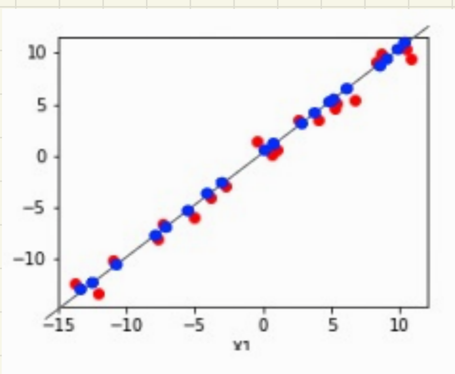
How to Find Best Line?



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

$$Var(D) = 109.45$$

$$\Downarrow$$



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

$$V(D_{PCA}) = 109.03$$

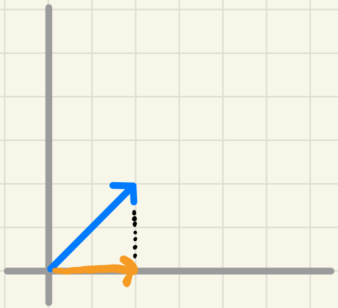
Linear Transformations: Projection onto Axis

$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $\quad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Consider

$$A_1 v = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In General: $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$



$$A_1 x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1(x_1) + 0(x_2) \\ 0(x_1) + 0(x_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$

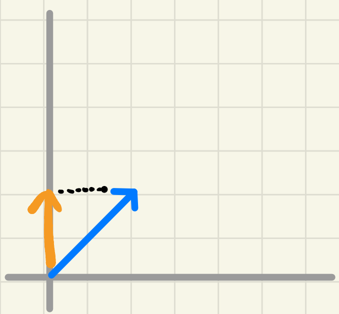
Linear Transformations: Projection onto Axis

$A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Consider

$$A_2 v = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

In General: $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$



$$A_2 x = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0(x_1) + 0(x_2) \\ 0(x_1) + 1(x_2) \end{pmatrix} = \begin{pmatrix} 0 \\ x_2 \end{pmatrix}$$

Linear Trans: Rotations

$$A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

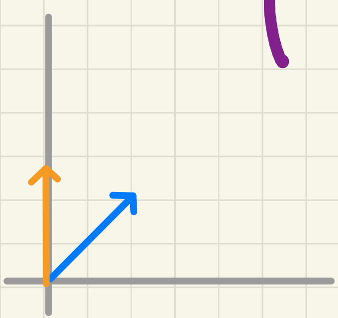
$$v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Consider $\theta = \pi/4$

In general, what does $A_\theta x$ do?

Rotate x CCW by θ

$$A_{\pi/4} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$



$$A_{\pi/4} v = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2}(1) - \frac{\sqrt{2}}{2}(1) \\ \frac{\sqrt{2}}{2}(1) + \frac{\sqrt{2}}{2}(1) \end{pmatrix} = \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix}$$

Linear Trans: Rotations

$$A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

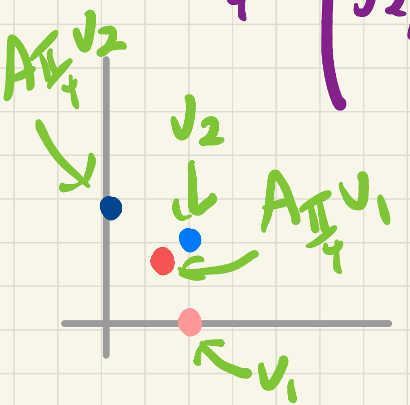
$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\quad v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Consider $\theta = \pi/4$

$$A_{\pi/4} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

In general, what does $A_\theta x$ do?

Rotate x CCW by θ



$$A_{\pi/4}[v_1 \; v_2] = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{array}{cc} v_1 & v_2 \\ \downarrow & \downarrow \\ \multicolumn{2}{c}{\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}} \end{array} = \begin{array}{cc} A_{\pi/4} v_1 & A_{\pi/4} v_2 \\ \downarrow & \swarrow \\ \multicolumn{2}{c}{\begin{pmatrix} 0 & \frac{\sqrt{2}}{2} \\ \sqrt{2} & \frac{\sqrt{2}}{2} \end{pmatrix}} \end{array}$$

Linear Trans: Rotations

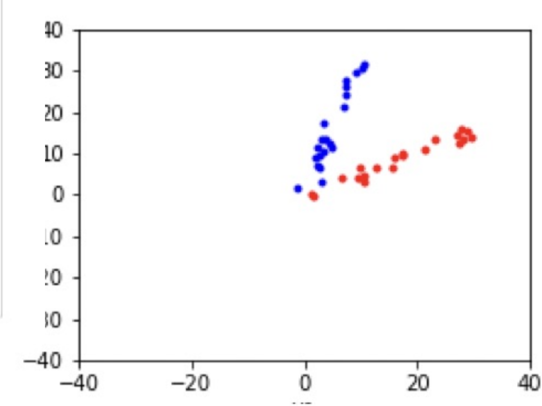
$$A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

How to rotate all of D by θ?

$$A_\theta D^T$$

Consider $\theta = \pi/4$ 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

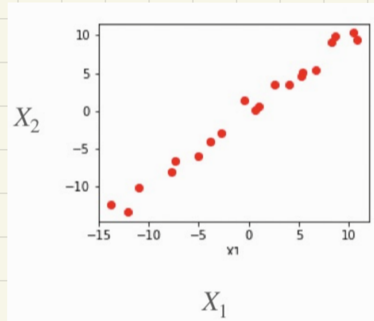


$$A_{\pi/4} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

$$A_{\pi/4} \begin{bmatrix} -13.75\ldots & 5.33\ldots & \ldots \\ -12.36\ldots & 5.03\ldots & \ldots \end{bmatrix} = \text{Blue points!}$$

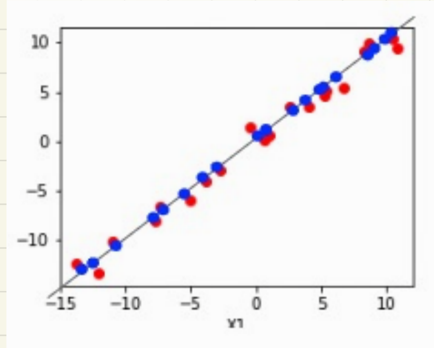
How does this relate to PCA?

We want to rotate the axis to better fit the data



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

$\Downarrow$ PCA



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

Linear Transformation: Scale

$$S = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad R = \begin{pmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{pmatrix} \quad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

What is SRv?

$$SRv = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$S(Rv) = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$