Multiple Linear Regression

Foundations of Data Analysis

March 15, 2023

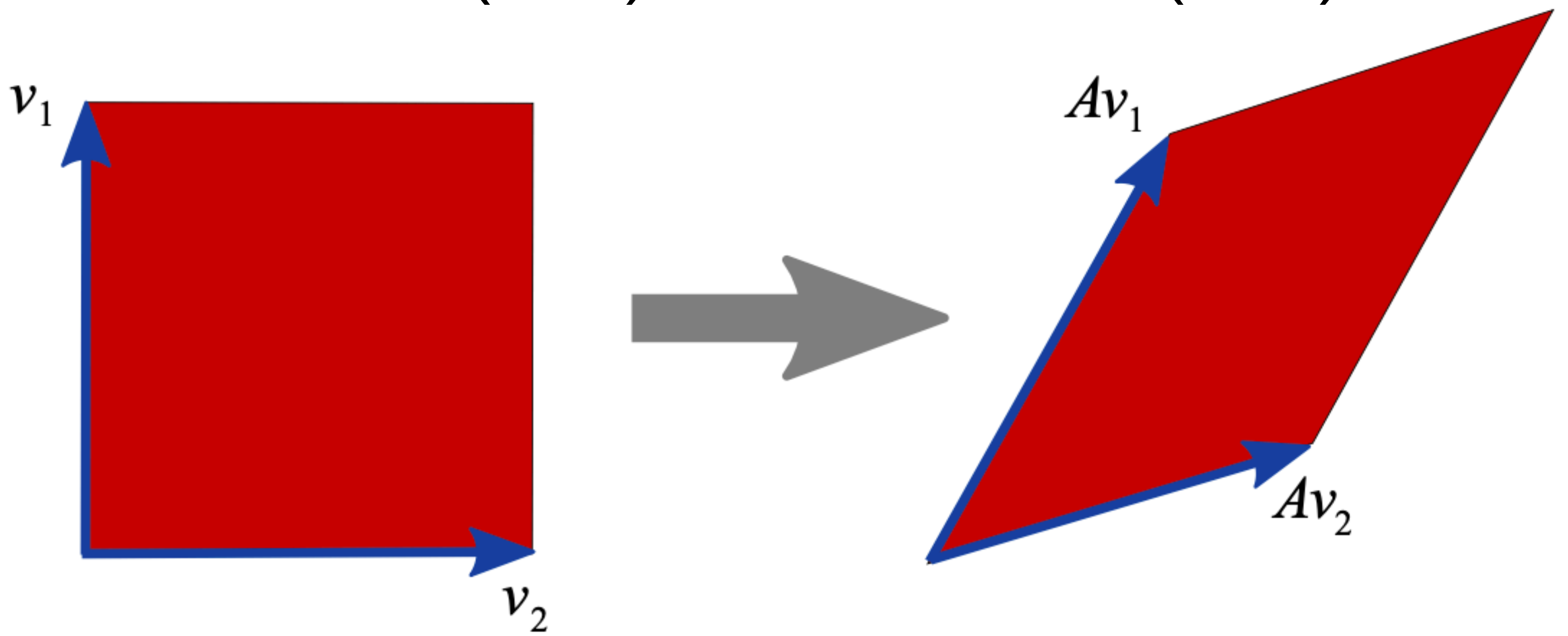
Matrices as Transformations

Consider a 2D matrix and coordinate vectors in $\mathbb{R}^2$:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then Av_1 and Av_2 result in the columns of A:

$$Av_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \qquad Av_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$



Transformation in Real-World Applications

- 3D face image matching/registration of smart phones

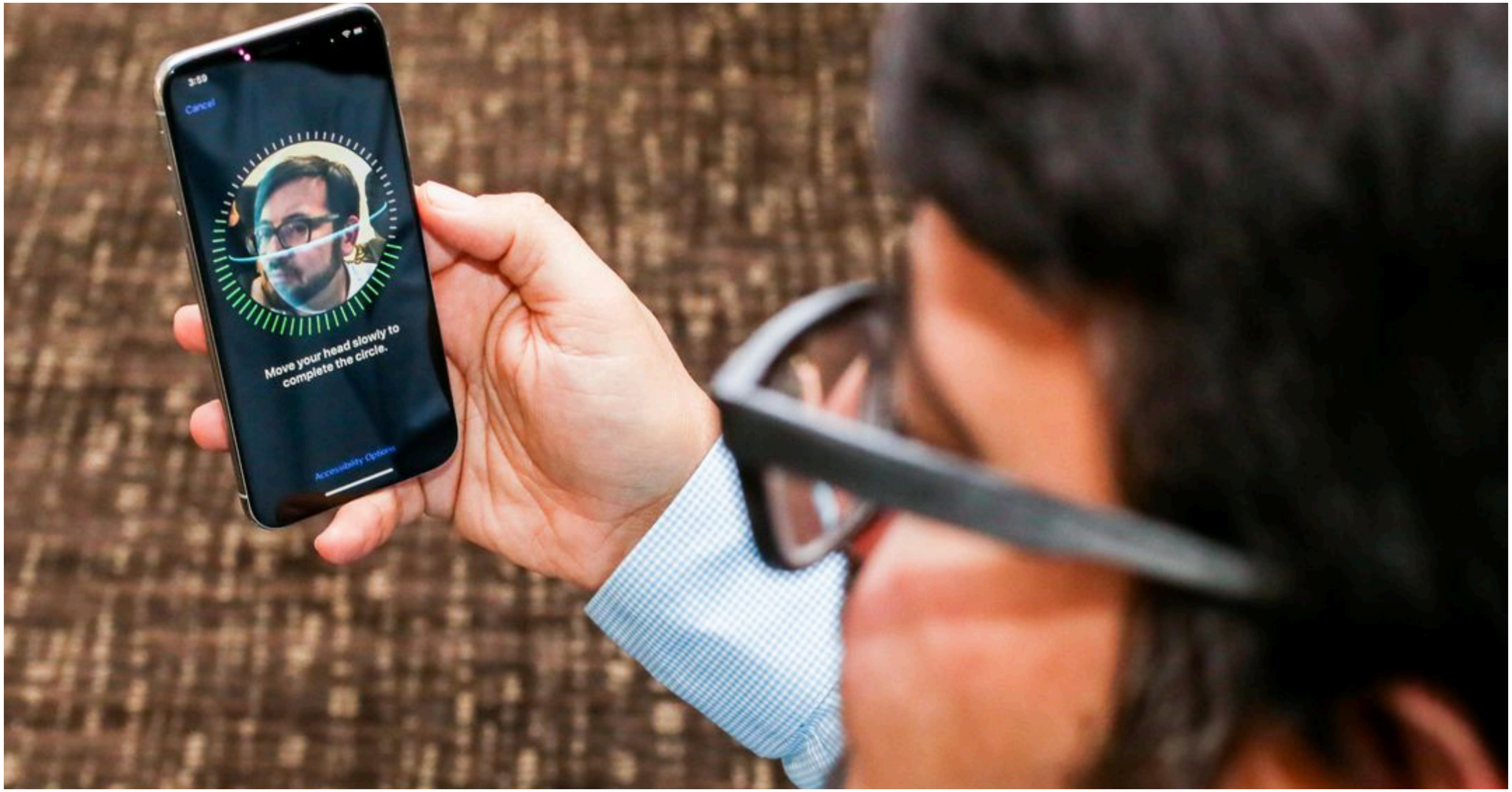
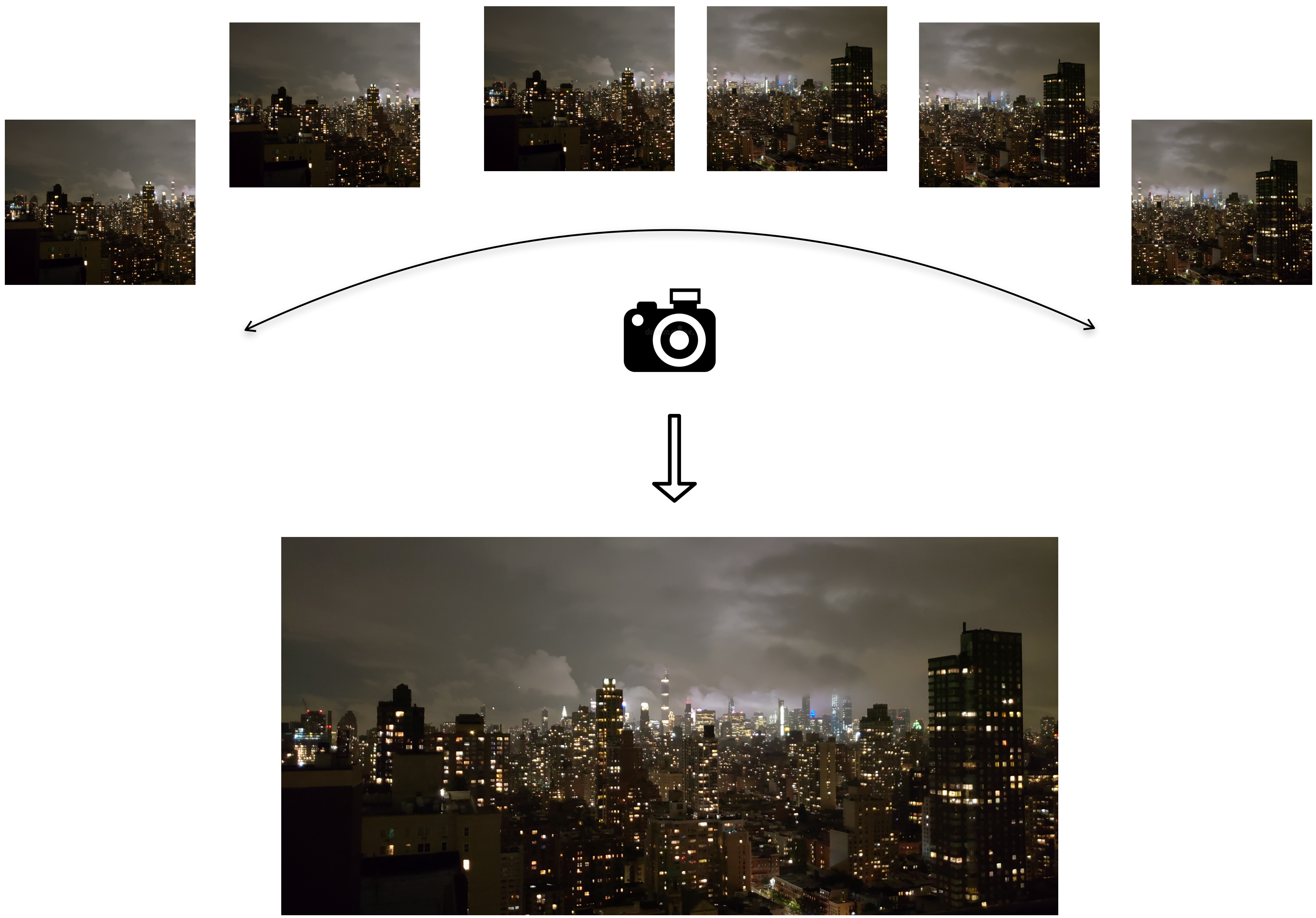


Image credits: Sarah Tew

Transformation in Real-World Applications

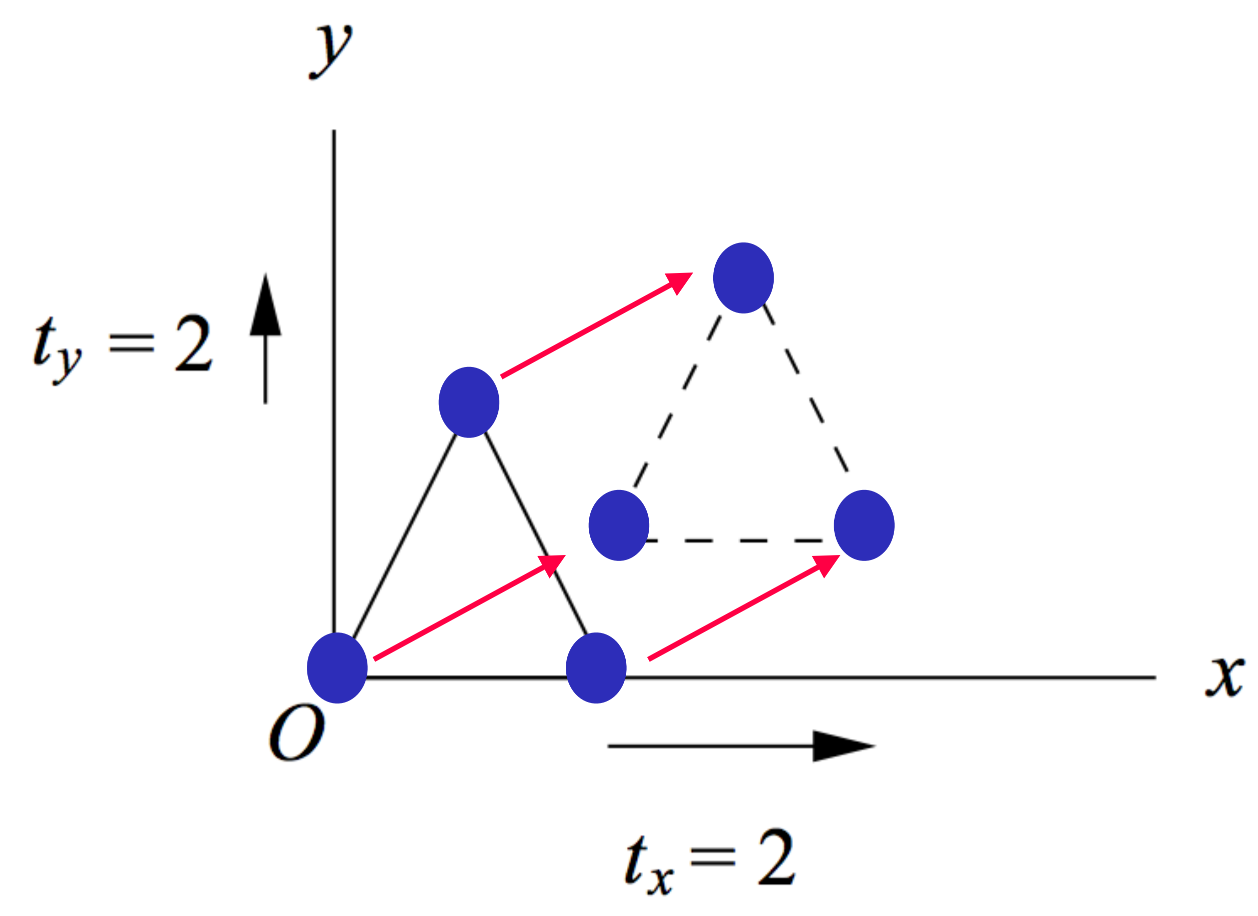
- Image stitching (panorama):



Transformation To Distort Images

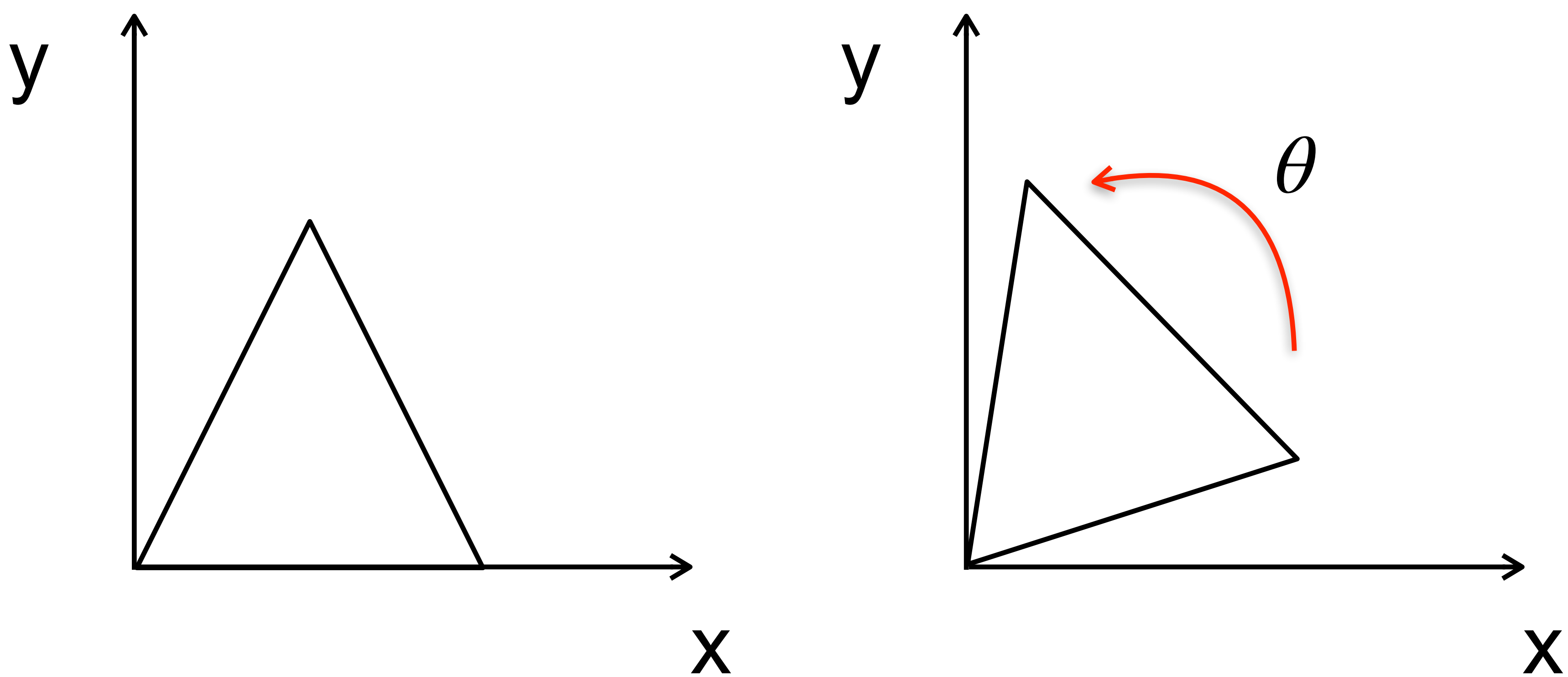
Translation Matrix

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix}$$



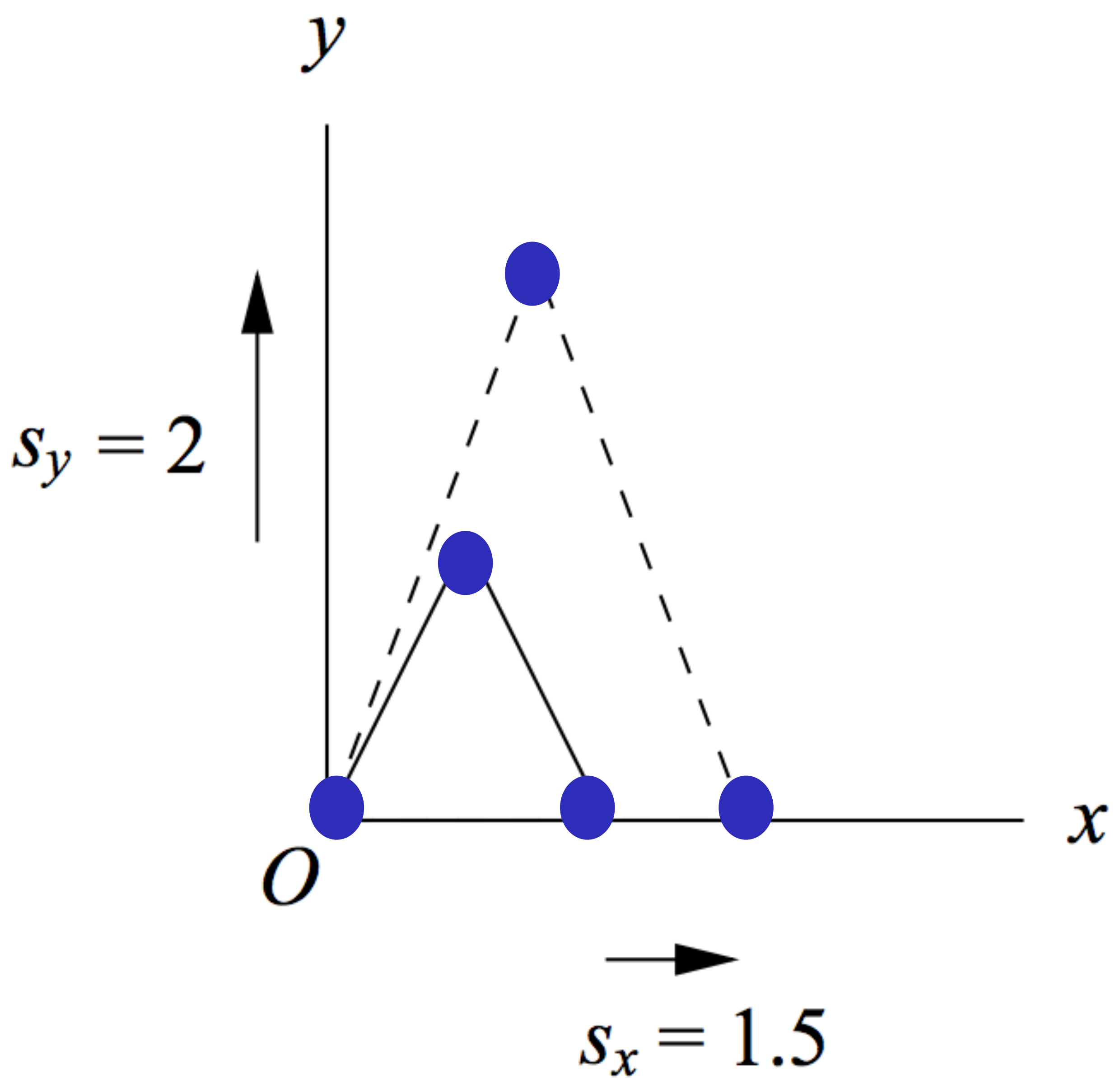
Rotation Matrix

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



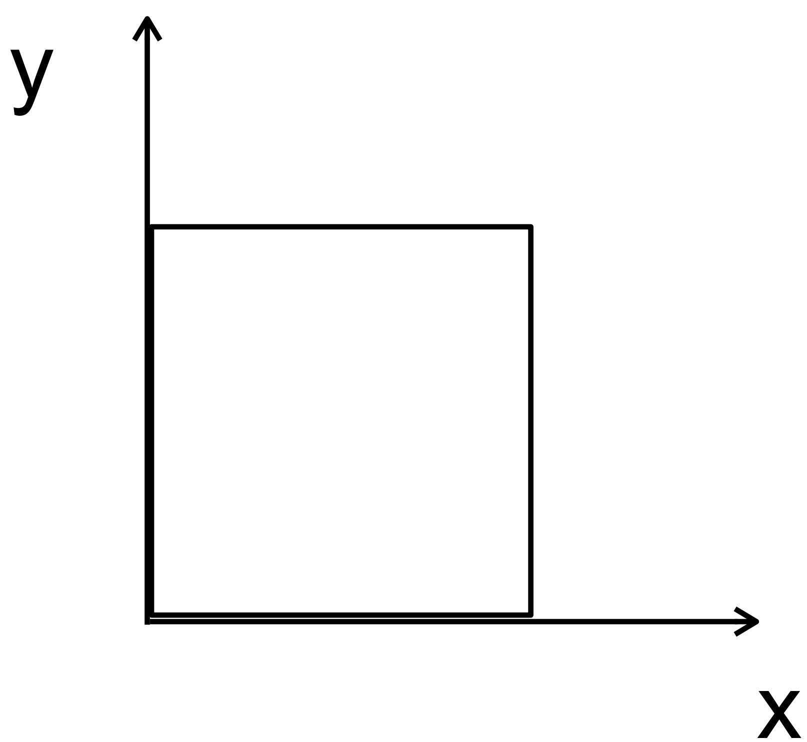
Scaling Matrix

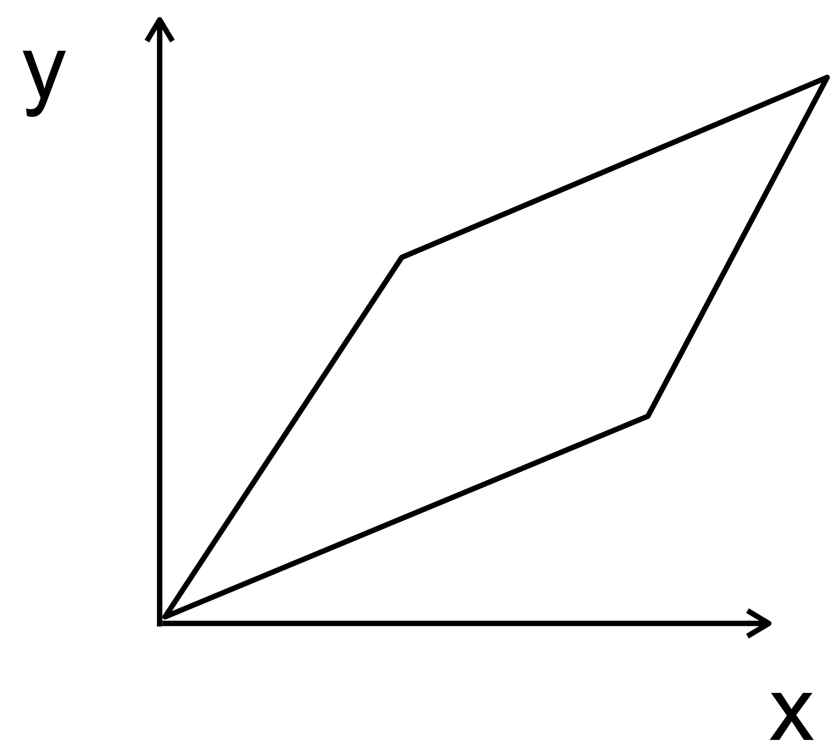
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



Shearing Matrix

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & A \\ B & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + Ay \\ xB + y \end{bmatrix}$$





Affine Transformation Matrix

- Affine transformation (translation, rotation, shearing, scaling)

$$T_{\text{affine}}(I) = R \cdot M \cdot I + t$$

I : original image coordinates R : rotation matrix

t : translation M: shearing + scaling matrix

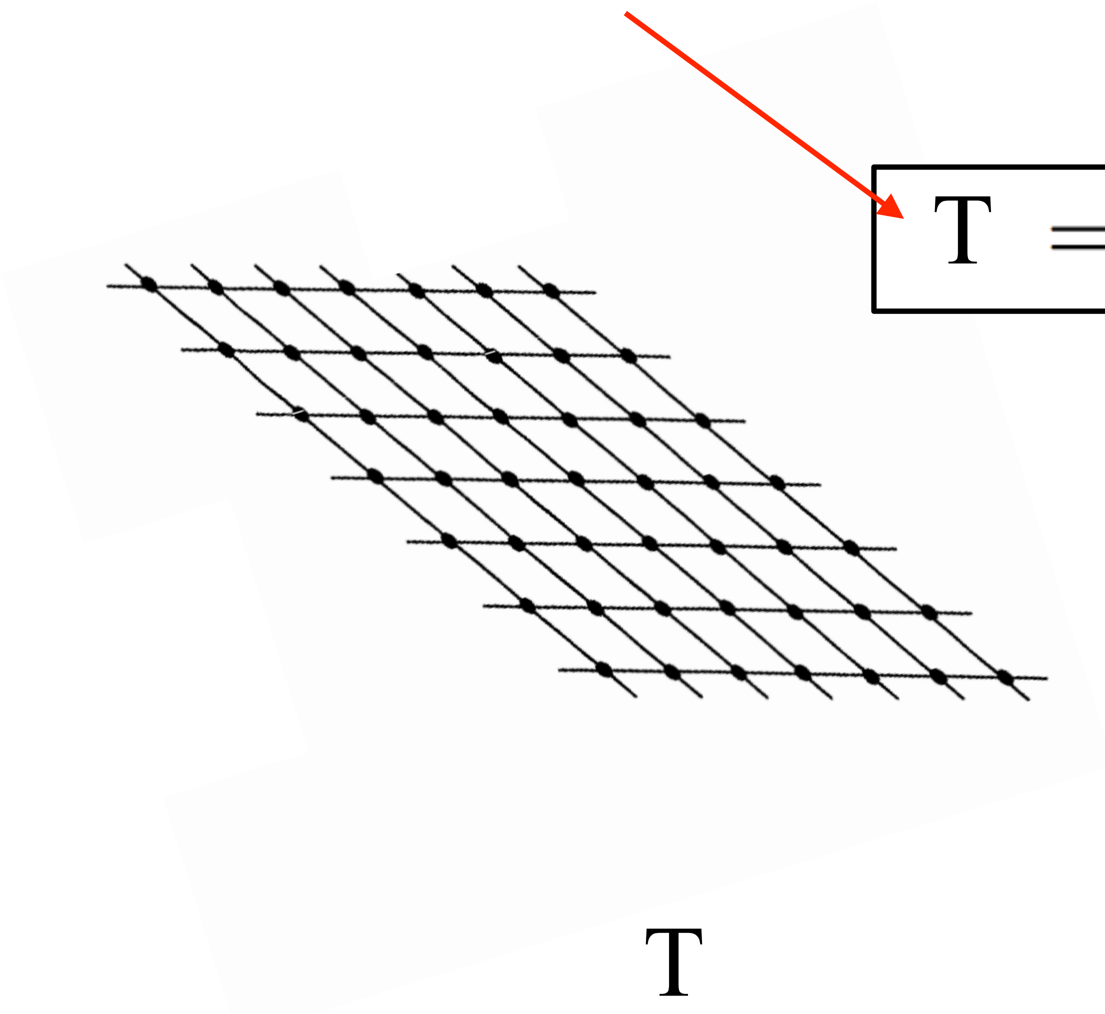
$$\begin{bmatrix} s_x(\cos\theta - B\sin\theta) & -s_y(\sin\theta + A\cos\theta) & t_x \\ s_x(\sin\theta + B\cos\theta) & s_y(\cos\theta + A\sin\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Affine Transformation Matrix

- Find affine transformation by solving linear equations

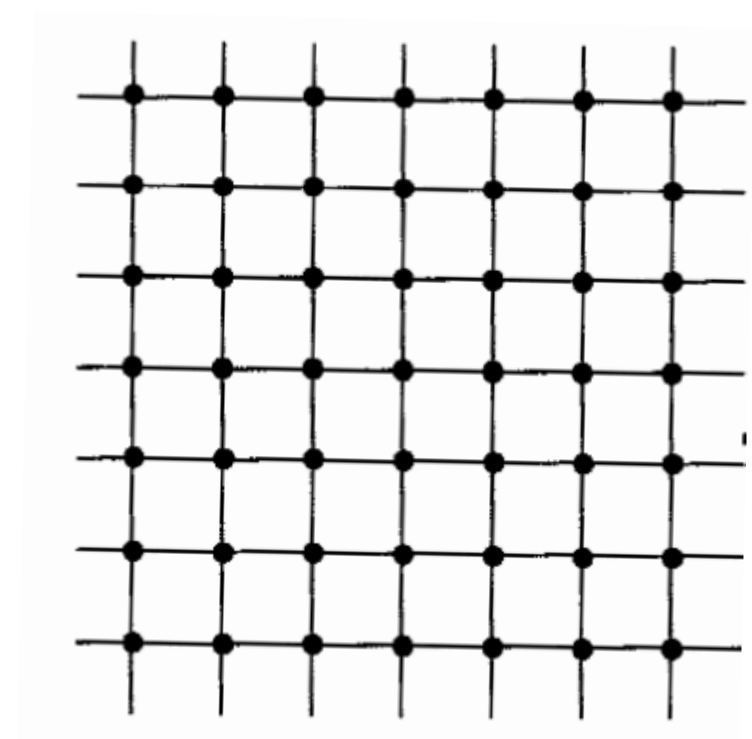
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\boxed{\mathrm{T} = \mathbf{A}\,\mathrm{S}}$$



T

= A *



S

Multilinear Regression: d features

What if x_i has d features $(x_{i1}, x_{i2}, \ldots x_{id})$?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{id}\beta_d + \epsilon_i$$

Written as matrix-vector multiplication:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1d} \\ x_{21} & x_{22} & \ldots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Or, in matrix notation:

$$y = \mathbf{X}\beta + \epsilon$$

Multilinear Regression: Adding an Intercept

Use β_1 for the intercept:

$$y_i = \beta_1 + x_{i2}\beta_2 + \ldots + x_{id}\beta_d + \epsilon_i$$

First column of X is ones:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \ldots & x_{1d} \\ 1 & x_{22} & \ldots & x_{2d} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \ldots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Or, in matrix notation:

$$y = \mathbf{X}\beta + \epsilon$$

Least Squares Problem

Regression equation:

$$y = X\beta + \epsilon$$

Minimize the sum-of-squared error:

$$\begin{aligned} \text{SSE}(\beta) &= \|\epsilon\|^2 \\ &= \|y - X\beta\|^2 \\ &= (y - X\beta)^T(y - X\beta) \end{aligned}$$

LS Solution

Derivative:

$$\frac{\partial}{\partial\beta}SSE(\beta) = -X^T(y - X\beta) = -X^Ty + X^TX\beta$$

$$X^TX\beta = X^Ty \qquad \text{set derivative to zero}$$

$$(X^TX)^{-1}(X^TX)\beta = (X^TX)^{-1}X^Ty \qquad \text{multiply by } (X^TX)^{-1}$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty \qquad \text{solve for } \beta$$

What is SVD?

Decompose a matrix A into three parts:

$$A = USV^T$$

The matrices U, S, and V have special properties

Why is SVD Useful?

Many applications in data analysis, including:

- Least squares fitting of data
- Dimensionality reduction
- Correlation analysis

When you need to solve the linear equation:

$$Y = Ax + b$$

Review: Data Tables

	ID	M.F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay	RightHippoVol	LeftHippoVol
0	OAS1_0002_MR1	F	R	55	4	1.0	29	0.0	1147	0.810	1.531	NaN	4230	3807
1	OAS1_0003_MR1	F	R	73	4	3.0	27	0.5	1454	0.708	1.207	NaN	2896	2801
2	OAS1_0010_MR1	M	R	74	5	2.0	30	0.0	1636	0.689	1.073	NaN	2832	2578
3	OAS1_0011_MR1	F	R	52	3	2.0	30	0.0	1321	0.827	1.329	NaN	3978	4080
4	OAS1_0013_MR1	F	R	81	5	2.0	30	0.0	1664	0.679	1.055	NaN	3557	3495
5	OAS1_0015_MR1	M	R	76	2	NaN	28	0.5	1738	0.719	1.010	NaN	3052	2770
6	OAS1_0016_MR1	M	R	82	2	4.0	27	0.5	1477	0.739	1.188	NaN	3421	3119
7	OAS1_0018_MR1	M	R	39	3	4.0	28	0.0	1636	0.813	1.073	NaN	4496	4283
8	OAS1_0019_MR1	F	R	89	5	1.0	30	0.0	1536	0.715	1.142	NaN	3760	3167
9	OAS1_0020_MR1	F	R	48	5	2.0	29	0.0	1326	0.785	1.323	NaN	3557	3394
10	OAS1_0021_MR1	F	R	80	3	3.0	23	0.5	1794	0.765	0.978	NaN	3715	3019
11	OAS1_0022_MR1	F	R	69	2	4.0	23	0.5	1447	0.757	1.213	NaN	3258	3566
12	OAS1_0023_MR1	M	R	82	2	3.0	27	0.5	1420	0.710	1.236	NaN	3217	2160
13	OAS1_0026_MR1	F	R	58	5	1.0	30	0.0	1235	0.820	1.421	NaN	3783	3535
14	OAS1_0028_MR1	F	R	86	2	4.0	27	1.0	1449	0.738	1.211	NaN	3452	3100
15	OAS1_0030_MR1	F	R	65	2	3.0	29	0.0	1392	0.764	1.261	NaN	3969	3406

Row: individual data point

Column: particular dimension or feature

Review: Matrices

A matrix is an $n \times d$ array of real numbers:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1d} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2d} \\ \vdots & & & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nd} \end{pmatrix}$$

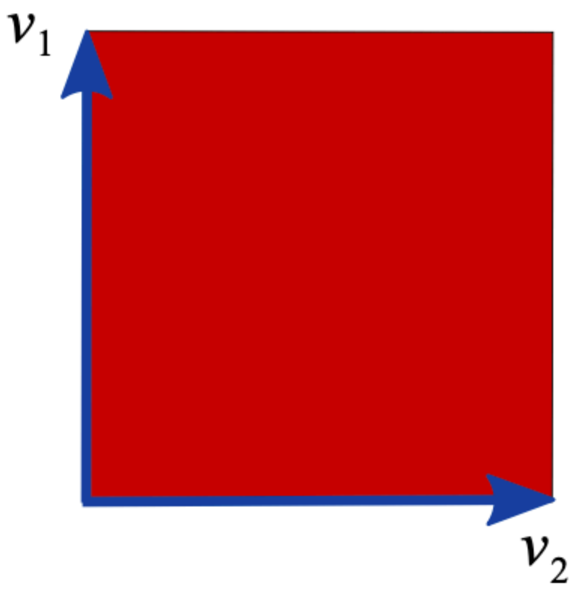
A **data matrix** is n data points, each with d features

Orthogonal Matrices

A matrix U is called **orthogonal** if the columns of U have unit length and are orthogonal to each other:

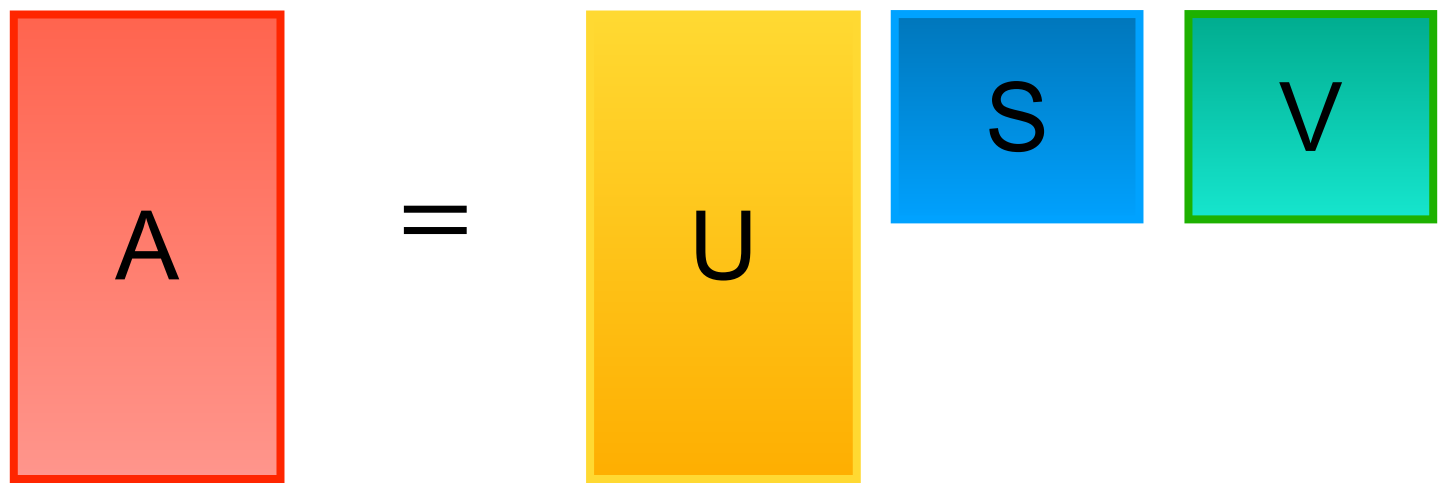
Unit length: $\|u_{\bullet i}\| = 1$

Orthogonal: $\langle u_{\bullet i}, u_{\bullet j} \rangle = 0$



SVD

$$A = USV^T$$



$U : n \times d$ orthogonal matrix (left singular vectors)

$S : d \times d$ diagonal matrix (singular values)

$V : d \times d$ orthogonal matrix (right singular vectors)

Application: Orthogonal Procrustes Analysis

Problem:

Find the rotation R^* that minimizes distance between two $d \times k$ matrices A, B:

$$R^* = \arg \min \| \mathrm{RA} - \mathrm{B} \|^2$$

Solution:

Let $U \Sigma V^T$ be the SVD of BA^T, then

$$R^* = UV^T$$