Linear Algebra: Singular Value Decomposition

Glenn Bruns

CSUMB

Much of the material in these slides comes from Géron's notes:
https://github.com/ageron/handson-ml/blob/master/math_linear_algebra.ipynb

Learning outcomes

After this lecture you should be able to:

1. Define:
   - matrix determinants
   - singular value decomposition (SVD)
   - eigenvalues and eigenvectors
2. Describe training in linear regression with matrices

Matrix inversion and determinants

Review of matrix inversion:

A square matrix A is invertible if there exists a square matrix B such that

$$AB = I$$

If A is invertible, then its inverse is unique, and written A^{-1}

A square matrix that is not invertible is called singular (or degenerate).

Determinants:

The determinant of a square matrix A, written $|A|$, is a scalar – a single number.

Key fact:

$|A| = 0$ iff A is singular

Defining determinant

The determinant of square matrix can be defined recursively:

1. if A is a 1×1 matrix: $|A| = A_{1,1}$
2. otherwise:

$$|A| = A_{1,1} \times |A^{(1,1)}| - A_{1,2} \times |A^{(1,2)}| + A_{3,1} \times |A^{(1,3)}| - A_{1,4} \times |A^{(4,1)}| + \cdots \pm A_{1,n} \times |A^{(1,n)}|$$

(where $A^{(i,j)}$ is matrix A without row i and column j).

Example: A = $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 0 \end{pmatrix}$ Compute a term for each column:

$|A| = 1 \times \left|\begin{pmatrix} 5 & 6 \\ 8 & 0 \end{pmatrix}\right| - 2 \times \left|\begin{pmatrix} 4 & 6 \\ 7 & 0 \end{pmatrix}\right| + 3 \times \left|\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}\right|$ Continuing:

$\left|\begin{pmatrix} 5 & 6 \\ 8 & 0 \end{pmatrix}\right| = 5 \times 0 - 6 \times 8 = -48$, etc. Result is 27, so A is _____?

Defining determinant

The determinant is simple for a 2x2 matrix:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \mathrm{ad} - \mathrm{bc}$$

Defining determinant

Interesting properties of determinants:

1. For the identity matrix, $|A| = 1$.

2. If A is a square matrix with two equal rows, or two equal columns, then $|A| = 0$

3. If you get matrix B by swapping two rows of a square matrix A, then

$$|B| = -|A|$$

Recall: matrices as transformers

Remember that a matrix A can be thought of as a function, or transformer, of a vector u. For example:

If $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ and $u = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$ then $Au = \begin{pmatrix} -2 \\ 1 \\ 9 \end{pmatrix}$

Also, you could apply the transformation A to multiple vectors at once:

$$B = \begin{pmatrix} -1 & 0 \\ 2 & 1 \\ 3 & -2 \end{pmatrix}$$

$$AB = \begin{pmatrix} -2 & 0 \\ 1 & ? \\ 9 & ? \end{pmatrix}$$

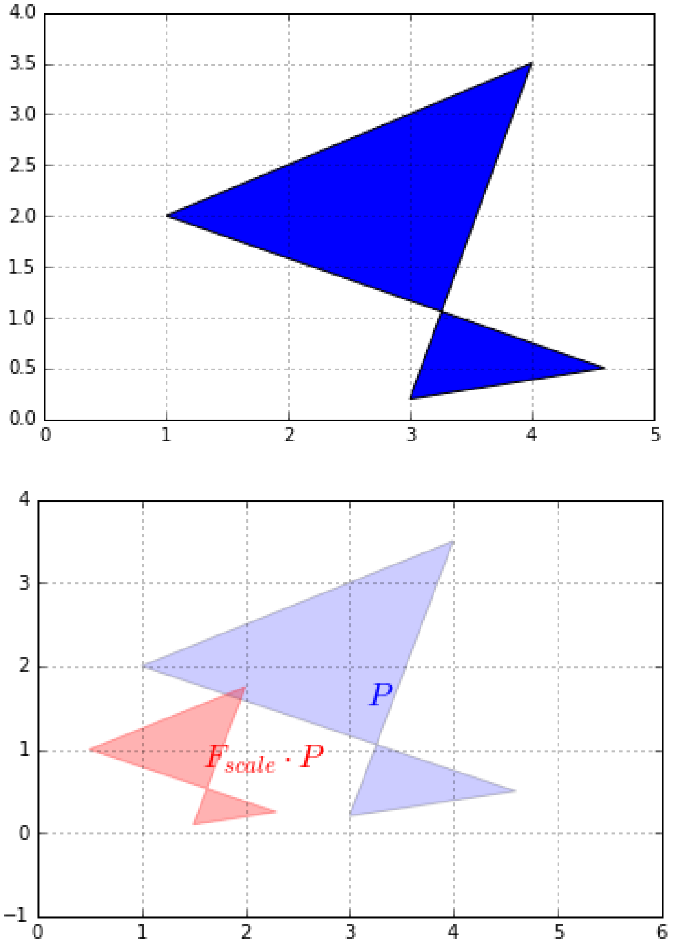
Determinants in NumPy

```
# four points
P = np.array([
        [3.0, 4.0, 1.0, 4.6],
        [0.2, 3.5, 2.0, 0.5]
    ])

# a transformation
F_scale = np.array([
        [0.5, 0],
        [0, 0.5]
    ])

# apply the transformation
P_scaled = F_scale.dot(P)

# what's the determinant of
F_scale?
LA.det(F_scale)
```



Result = 0.25, So F_scale is invertible: we can "undo" the scaling

figures: Géron text

Composing transformations

An example of a matrix transforming a vector:

If $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ and $u = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$ then $Au = \begin{pmatrix} -2 \\ 1 \\ 9 \end{pmatrix}$

Applying multiple transformations might look like this:

$A(B(Cu))$ which equals $(ABC)u$ (why?)

In code, can turn three transformations into a single one.

Instead of:

```
P_squeezed_then_sheared = F_shear.dot(F_squeeze.dot(P))
```

write:

```
F_squeeze_then_shear = F_shear.dot(F_squeeze)
P_squeezed_then_sheared = F_squeeze_then_shear.dot(P)
```

Singular value decomposition (SVD)

Integers can be “decomposed” into a product of primes.

Matrices can be “decomposed” into the product of three simple matrices.

Any $m \times n$ matrix A can be decomposed like this:

$$A = U\Sigma V^T$$

where

- U is a rotation matrix (an $m \times m$ orthogonal matrix)
- Σ is a scaling & projecting matrix (an $m \times n$ diagonal matrix)
- V^T is a rotation matrix (an $n \times n$ orthogonal matrix)

A square matrix H is orthogonal if its inverse is the same as its transpose:

$$H^{-1} = H^T$$

as a result:

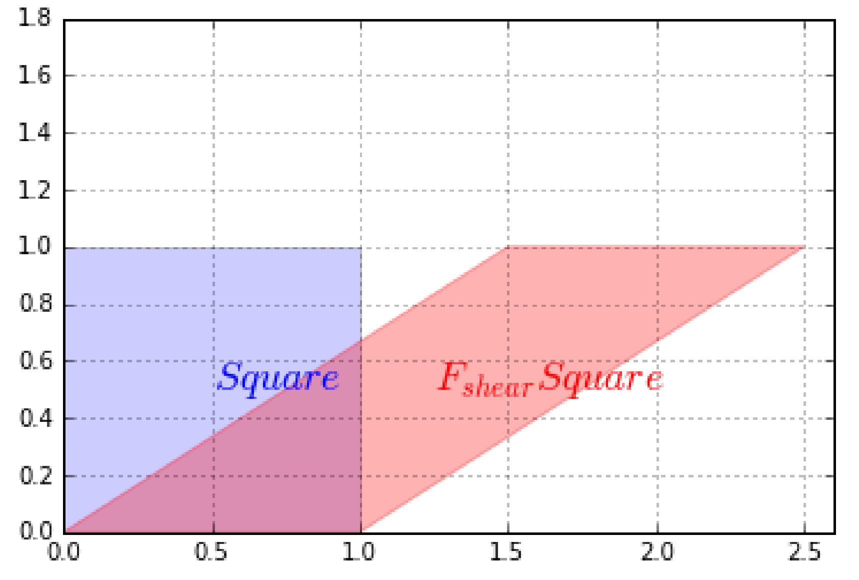
$$HH^T = H^T H = 1$$

SVD Example, part 1

```
# four corners of a square
Square = np.array([
        [0, 0, 1, 1],
        [0, 1, 1, 0]
])

# a "shearing" transformation
F_shear = np.array([
        [1, 1.5],
        [0, 1]
    ])

# apply the transformation
F_shear.dot(Square)
```



figures: Géron text

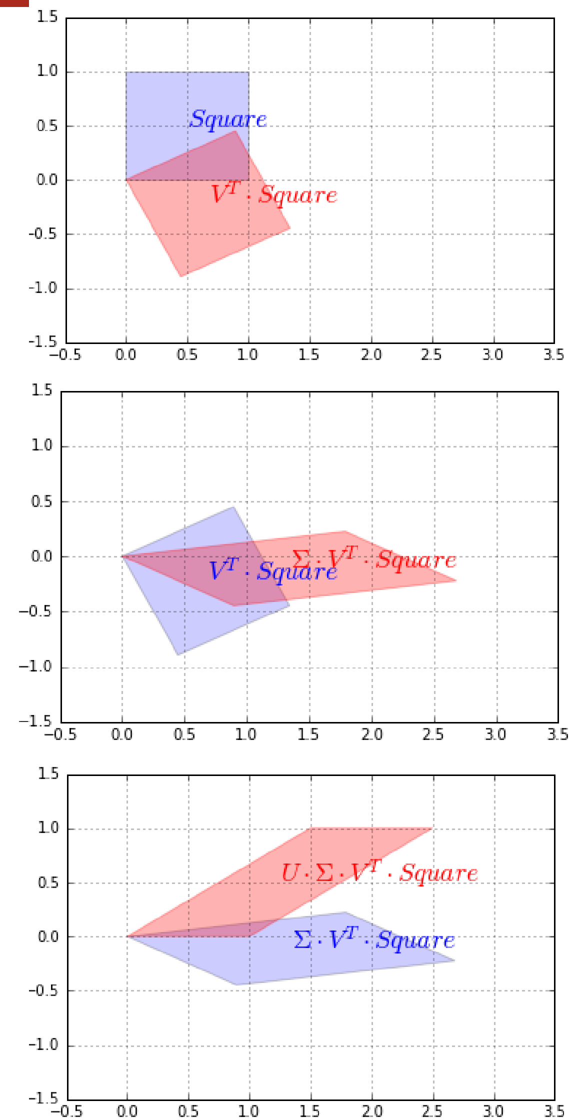
SVD Example, part 2

```
# decompose F_shear
U, S_diag, V_T = LA.svd(F_shear)
U
array([[ 0.89442719, -0.4472136 ],
       [ 0.4472136 ,  0.89442719]])

np.diag(S_diag)
array([[ 2. ,  0. ],
       [ 0. ,  0.5]])

# put it back together
U.dot(np.diag(S_diag)).dot(V_T)
array([[ 1. ,  1.5],
       [ 0. ,  1. ]])

F_shear
array([[ 1. ,  1.5],
       [ 0. ,  1. ]])
```



rotate

scale

rotate

figures: Géron text

Eigenvalues and Eigenvectors

These sounds exotic and difficult but they're not.

Warmup: Let matrix A be

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Is there a 2 x 2 matrix B such that

$$AB = B$$

?

A value x such that

$$f(x) = x$$

is called a fixed point of f.

Some functions have no fixed points, some functions have many fixed points.

Eigenvalues and Eigenvectors

Suppose we "apply" a matrix A to a vector v:

$$f(v) = Av$$

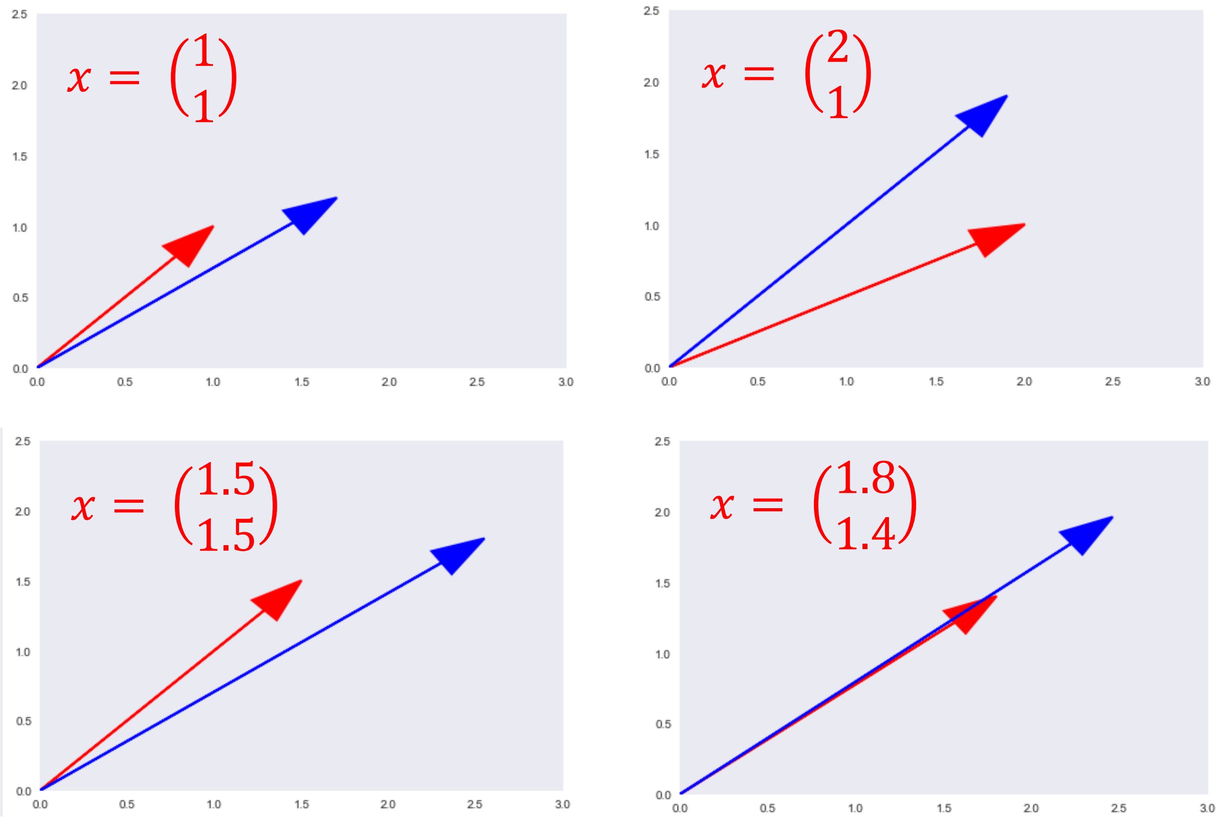
Suppose the output vector Av points in the same direction as v, but might be scaled differently.

Eigenvalues and Eigenvectors

$$A = \begin{pmatrix} 0.2 & 1.5 \\ 0.7 & 0.5 \end{pmatrix}$$

What is $A\begin{pmatrix} 1 \\ 1 \end{pmatrix}$?

Plot x and Ax for different vectors x:



Eigenvalues and Eigenvectors

Suppose we “apply” a matrix A to a vector v:

$$f(v) = Av$$

Suppose the output vector Av points in the same direction as v, but might be scaled differently.

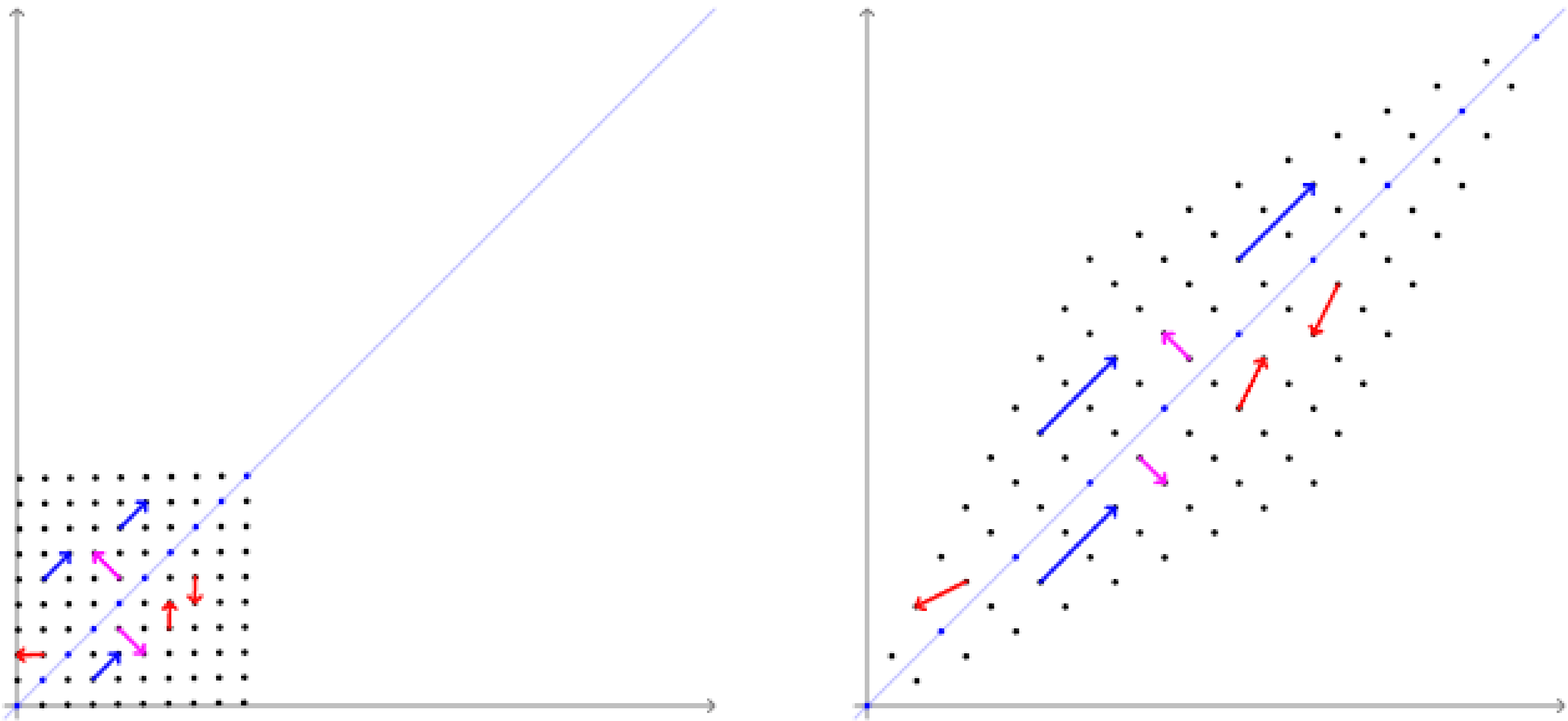
In other words:

$$Av = cv \qquad \textit{(for some scalar c)}$$

If this is true, we say:

- v is an eigenvector of A
- c is the eigenvalue associated with v

Another visualization of Eigenvectors



On the left we see a bunch of vectors. On the right we see the vectors after being transformed by the square matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. The blue and purple vectors are eigenvectors of A, because their directions are not changed by the transformation.

(This example paraphrased from the Wikipedia entry 'Eigenvalues and Eigenvectors')

Training in linear regression

Earlier we said we want to choose β to minimize this:

$$RSS(\beta) = (y - X\beta)^T (y - X\beta)$$

X is an $n \times (p+1)$ matrix (each row is a feature vector)

y is an n-vector of labels

We want to find the value of β that minimizes $RSS(\beta)$.

To do so, differentiate it with respect to β, giving:

$$X^T(y - X\beta) = 0$$

If $X^T X$ is non-singular, then the unique solution is:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

(A variable with a "hat", like $\hat{\beta}$, is often used to signify the "estimated value of". In this case, the estimated value of β)

Material on this slide based on the presentation in *The Elements of Statistical Learning* by Hastie et al

Summary

1. Determinants
2. Singular Value Decomposition (SVD)
3. Eigenvalues and Eigenvectors

The last two topics are used in dimensionality reduction, and other topics.