Linear Algebra

Matrices

$$m \times n \text{ matrix } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \text{ the "transpose" } A^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ a_{12} & & \\ \vdots & & \\ a_{1m} & & \end{pmatrix},$$

where the transpose interchanges columns and rows. If $m = n$, then A is called "square".

Matrix multiplication

If A is a $m \times n$ matrix and B is a $n \times l$ matrix, then AB is a $m \times l$ matrix where

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

We can think of this as dot products of vectors:

$$AB = \begin{pmatrix} \vec{a}_1 \rightarrow \\ \vec{a}_2 \rightarrow \\ \vdots \\ \vec{a}_m \rightarrow \end{pmatrix} \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_l \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

$$(AB)_{ij} = \vec{a}_i \cdot \vec{b}_j = \vec{a}^T \vec{b}.$$

Identity matrix

$$\text{square matrix } I = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & 0 & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & 0 & & & & 1 & \\ & & & & & & 1 \end{pmatrix}.$$

I is the "identity matrix" because

$$AI = A$$
$$IA = A,$$

for all matrices A.

Inverse of a square matrix

$$AA^{-1} = A^{-1}A = I.$$

Note that $(A^{-1})^{-1} = A$. Finding the inverse of a matrix is slow in practice. In general:

$$A^{-1} = \frac{1}{|A|}(C_{ij})^T = \frac{1}{|A|} \begin{pmatrix} C_{11} & C_{21} & \cdots \\ C_{12} & \ddots & \\ \vdots & & C_{ji} \end{pmatrix},$$

where $|A| = \det A$. Therefore A^{-1} is one over the determinant of A times the transpose of the cofactor matrix (aka the "adjoint of A").

$$C_{ij} = (-1)^{i+j} M_{ij},$$

where the minor M_{ij} is the determinant of the submatrix obtained from A by deleting row i and column j.

Determinants of square matrices

Eg. 2×2 matrix:

$$det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

The determinant of larger matrices can be defined by induction:

$$|A| = \det A = \sum_{j=1}^{n} a_{ij} (-1)^{i+j} M_{ij},$$

where as above M_{ij} is the determinant of the submatrix formed by A without row i and column j:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & \big(\text{submatrix "11"}\big) & \end{pmatrix}, \text{ so that } |A| = a_{11} M_{11} - a_{12} M_{12} + \ldots$$

Eg. 3×3 matrix:

$$\det \begin{pmatrix} i & j & k \\ a & b & c \\ d & e & f \end{pmatrix} = i(bf - ce) - j(af - cd) + k(ae - bd).$$

Eg. Inverse of a 2×2 matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\begin{aligned} A^{-1} &= \frac{1}{|A|} \begin{pmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{pmatrix} = \frac{1}{|A|} \begin{pmatrix} M_{11} & -M_{21} \\ -M_{12} & M_{22} \end{pmatrix} \\ &= \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \end{aligned}$$

Check:

$$\begin{aligned} A^{-1} A &= \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \frac{1}{ad - bc} \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \qquad \checkmark \end{aligned}$$

Linear independence

$$c_1 \vec{v}_1 + c_2 \vec{v}_2 + c_3 \vec{v}_3 + \cdots + c_n \vec{v}_n = 0 \Leftrightarrow c_1 = c_2 = \cdots = c_n = 0.$$

If n vectors are not linearly independent, then they span a subspace of dimension $< n$. Eg. Two vectors $\vec{v}_1$ and $\vec{v}_2$ are linearly dependent if and only if they are multiples of each other: $\vec{v}_1 = a\vec{v}_2$. In this case $\vec{v}_1 + \vec{v}_2$ only span a 1 dimensional subspace.

Figure 1: Linearly dependent vectors

Orthogonality

Two vectors are orthogonal if $\vec{x} \cdot \vec{y} = 0$.

$$\vec{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \vec{y} = \begin{pmatrix} 4 \\ -2 \end{pmatrix}: \qquad \vec{x} \cdot \vec{y} = 1 \cdot 4 + 2(-2) = 0.$$

Orthogonal vectors are "perpendicular".

Figure 2: Orthogonal vectors

Eigenvalues and eigenvectors of symmetric matrices

Eigen means characteristic, so these are also sometimes called "characteristic" values and vectors. Eigenvalues and vectors are defined and related by the equation

$$A\vec{v} = \lambda\vec{v},$$

where $\vec{v}$ is an eigenvector and λ is an eigenvalue. (A consequence of this definition is that if $\vec{v}$ is an eigenvector, then so is $c\vec{v}$.)

To find eigenvalues, we use the theorem

$$(A - \lambda I)\vec{v} = 0 \Leftrightarrow \det(A - \lambda I) = 0.$$

This follows from a fact which is good to know: $\det X = \prod_i \lambda_i$. Similarly, $\operatorname{tr} X = \sum_i \lambda_i$.

We can therefore find eigenvalues by solving

$$\det(A - \lambda I) = \left| \begin{pmatrix} a_{11} - \lambda & a_{12} & & \cdots \\ a_{21} & a_{22} - \lambda & & \\ & \vdots & \ddots & \\ & & & a_{nn} - \lambda \end{pmatrix} \right| = 0,$$

which is an n^{th} order polynomial in λ, with n "roots", i.e. solutions for λ (these are either real or complex conjugate pairs).

Once a λ_i is known, $A\vec{v} = \lambda_i\vec{v}$ gives a set of linear equations which can be solved for the associated eigenvector $\vec{v}_i$.

Diagonalization

If A is an $n \times n$ matrix with n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and associated eigenvectors $\vec{v}_i$, then let

$$U = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix},$$

where the eigenvectors $\vec{v}$ are the columns of U. Now

$$\begin{aligned} AU &= A(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) \\ &= (A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_n) \\ &= (\lambda_1\vec{v}_1, \lambda_2\vec{v}_2, \ldots, \lambda_n\vec{v}_n) \\ &= U \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} = UD, \end{aligned}$$

where D is the diagonal matrix of eigenvalues. Manipulating the expression $AU = UD$ yields the identity

$$A = UDU^{-1}.$$

This is the eigen decomposition theorem, and it is an example of a similarity transform.

When A is a symmetric matrix, i.e., when $A = A^T$, then it is possible to make the $\vec{v}$s orthonormal, in which case

$$U^{-1} = U^T,$$

so that

$$A = UDU^T.$$

PCA is equivalent to diagonalization of a particular symmetric matrix — the covariance matrix.

The eigendecomposition allows some nice results:

$$A^2 = AA = (UDU^{-1})(UDU^{-1}) = UDDU^{-1} = UD^2U^{-1} = U \begin{pmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_n^2 \end{pmatrix} U^{-1},$$

and in the same way

$$A^m = UD^mU^{-1} = U \begin{pmatrix} \lambda_1^m & & & \\ & \lambda_2^m & & \\ & & \ddots & \\ & & & \lambda_n^m \end{pmatrix} U^{-1}.$$

So if we known the eigenvalues of A, we immediately know the eigenvalues of A^m. In particular,

$$A^{-1} = (UDU^{-1})^{-1} = (U^{-1})^{-1}D^{-1}U^{-1} = UD^{-1}U^{-1}.$$

It is easy to check that

$$A^{-1}A = (UD^{-1}U^{-1})(UDU^{-1}) = UD^{-1}DU^{-1} = UU^{-1} = I \qquad \checkmark$$

Therefore

$$A^{-1} = U \begin{pmatrix} 1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & \ddots & \\ & & & 1/\lambda_n \end{pmatrix} U^{-1}.$$