Matrix Decompositions

Trace of a matrix

It's the sum of diagonal elements of a matrix and is the only function satisfying these 4 properties :

- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- $\text{tr}(a\mathbf{A}) = a\,\text{tr}(\mathbf{A})$
- $\text{tr}(\mathbf{I}_n) = n$
- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$

Using the last property, we can get $\text{tr}(\mathbf{R^{-1}AR}) = \text{tr}(\mathbf{A})$, which means that similar matrices have the same trace, and thus for any linear map $f : V \to V$, no matter in what bases it's expressed, as long as it has same bases on either side, (before the map and after the map), we get the same trace for the matrix representing the map.

Eigenspace and geometric multiplicity of eigenvalues

For a matrix $\mathbf{A}$, the Eigenspace of the eigenvalue λ , denoted by E_λ is the null space of $\mathbf{A} - \lambda \mathbf{I}$.

The geometric multiplicity is the dimension of this null space and the algebraic multiplicity is the number of times this eigenvalue repeats in the roots of characteristic polynomial.

Eigenvalue properties

- A matrix $\mathbf{A}$ and its transpose $\mathbf{A}^T$ have the same eigenvalues (not the same eighenvectors) .
- Any linear map always has the same eigevalues, no matter how it is expressed , because the diagonal matrix in the decomposition stays the same (upto rearrangement)

- The sum of eigenvalues is just the trace and the product is the determinant (due to the characteristic polynomial) .

- Eigenvectors with different eigenvalues are linearly independent. Proof by induction : Suppose that this statement is true for k eigenvectors, then consider any $k + 1$ eigenvectors with distinct eigenvalues. Assume they are linearly independent, then we can write $0 = a_1\mathbf{v_1} + a_2\mathbf{v_2} + \ldots a_k\mathbf{v_k} + a_{k+1}\mathbf{v_{k+1}}$ where at least 2 of the coefficients must be non zero. Applying the matrix on both sides, we get $0 = a_1\lambda_1\mathbf{v_1} + \ldots a_k\lambda_k\mathbf{v_k} + a_{k+1}\lambda_{k+1}\mathbf{v_{k+1}}$. Now subtracting λ_{k+1} times the first equation from this, we get : $0 = a_1(\lambda_1 - \lambda_{k+1})\mathbf{v_1} + \ldots a_k(\lambda_k - \lambda_{k+1})\mathbf{v_k}$, which is basically screaming that k eigenvectors are linearly dependent since not all of the coefficients are 0 , which is a contradiction. And thus, we have extended the hypothesis to $k + 1$ eigenvectors. The base case is for 2 eigenvectors and is trivially true.

Positive definite matrices

Any matrix $\mathbf{A}$ such that $\mathbf{x^T A x} > 0 \ \ \forall\ \mathbf{x} \in V - \{0\}$ where V is a vector space is positive definite.

Positive definite matrices always have positive eigenvalues if they are real, since for any such matrix $\mathbf{A}$ with any eigenvector $\mathbf{v}$ with eigenvalue λ , we have $\mathbf{v^T A v} > 0 \implies \mathbf{v^T}\lambda\mathbf{v} = \lambda||\mathbf{v}|| > 0 \implies \lambda > 0$.

Also, by putting the basis vectors instead of $\mathbf{v}$, we get that the diagonal entries must also be positive.

Although positive definite matrices don't have to be symmetric, a lot of texts put the restriction in the definition itself. And I'm going to do the same. So when I say "positive definite", I actually mean "symmetric positive definite". Am I being a little bitch here ? Absolutely ! But you'll just have to deal with it :)

Symmetric matrices

Any matrix which is equal to its transpose is symmetric.

Spectral theorem : if $\mathbf{A}$ is a symmetric matrix, then the eigenvectors can be orthonormal and span the whole column space of $\mathbf{A}$.

Proof of orthogonality : Suppose $\mathbf{v_1}, \mathbf{v_2}$ are eigenvectotrs with unequal eigenvalues λ_1, λ_2 , then $\mathbf{v_1^T A^T} = (\mathbf{A v_1})^T = \lambda_1 \mathbf{v_1}^T \implies \mathbf{v_1^T A^T v_2} = \lambda_1 \mathbf{v_1^T v_2}$. But $\mathbf{A^T} = \mathbf{A}$. Thus, $\lambda_1 \mathbf{v_1^T v_2} = \mathbf{v_1^T A v_2} = \mathbf{v_1^T}(\lambda_2 \mathbf{v_2})$. But since $\lambda_1 \neq \lambda_2$, so we have $\mathbf{v_1^T v_2} = \mathbf{0}$ and thus $\mathbf{v_1}, \mathbf{v_2}$ are perpendicular if they have unequal eigenvalues. In other words, any two eigenvectors from eigenspaces of different eigenvalues are orthogonal. And we already know, that the eigenvectors in the same eigenspace can be chosen such that they are orthogonal. Thus we can create an orthonormal basis using all the eigenvectors.

The other fact that these eigenvectors span the column space is hard to prove. This fact allows us to write $\mathbf{A} = \mathbf{R D R^T} = \sum \lambda_i \mathbf{v_i v_i^T}$ where $\mathbf{D}, \mathbf{R}$ are a diagonal matrix and an orthogonal matrix (rectangular or not), and $\mathbf{v_i}, \lambda_i$ are the eigenvectors and corresponding eigenvalues for any symmetric matrix $\mathbf{A}$. If $\mathbf{A}$ is not symmetric , then $\mathbf{R}$ is not an orthogonal matrix, but still has the linearly independent eigenvectors of $\mathbf{A}$ as its column vectors, and the equation is still correct. This is called eigenvalue decomposition.

Producing symmetric positive semidefinite matrices

For any matrix $\mathbf{M}$ (rectangular, singular, anything..), the matrix $\mathbf{M^T M}$ is symmetric positive semi-definite because 1) ..$(\mathbf{M^T M})^T = \mathbf{M^T M}$ and 2) ....
$\mathbf{x^T M^T M x} = (\mathbf{M x})^T (\mathbf{M x}) = ||\mathbf{M x}||^2 \geq 0$. If the matrix is also non-singular, then the last inequality becomes strict and our composed matrix becomes symmetric definite.

Now you might wonder if the reverse can also be done ? That is, get a matrix $\mathbf{M}$, given the matrix

Now you might wonder if the reverse can also be done ? That is, get a matrix $\mathbf{M}$, given the matrix $\mathbf{M^TM}$. It is certainly possible and can be done by the Cholesky Factorization

Cholesky Factorization

Given a symmetric matrix $\mathbf{A}$ with rank n, we can write $\mathbf{A = LL^T}$ where $\mathbf{L}$ is an upper triangular matrix with n rows and positive diagonal entries. On paper, this would look like this :

$$
\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}
$$

The first column vector on the RHS has the elements $l_{11}^2, l_{11}l_{21}, \ldots l_{11}l_{n1}$ and thus first gives the value of l_{11} and then all of l_{i1}. Now that you know the first column, you can easily form the equations for $l_{22}l_{i2}$. Again, get l_{22} first and then get all the other values. Keep repeating this procedure and you'll end up knowing all columns of $\mathbf{L}^T$.

Singular Value decomposition

For any matrix $\mathbf{A}$, we can write $\mathbf{A = U\Sigma V^T} = \sum \sigma_i \mathbf{u_i v_i^T}$ where $\mathbf{U, V}$ are orthogonal square (rotation) matrices with unit column vectors $\mathbf{u_i, v_i}$ and $\mathbf{\Sigma}$ is a diagonal matrix (with 0 padding where necessary) with entries σ_i. On paper it looks like this :

We can get $\mathbf{U, V}$ by eliminating the other matrix like this :

$$\mathbf{A^TA} = \mathbf{V\Sigma U^TU\Sigma V^T} = \mathbf{V\Sigma^2 V^T}$$
$$\mathbf{AA^T} = \mathbf{U\Sigma V^TV\Sigma U^T} = \mathbf{U\Sigma^2 U^T}$$

and then performing eigenvalue decomposition of the matrices on LHS and **assigning** values to matrices. Note that the shape of $\mathbf{\Sigma^2}$ is possibly different in both equations. this is accounted for by some entries in the bigger version being 0.

Notice that although it looks like it, $\mathbf{\Sigma}$ is not the diagonal matrix obtained on eigenvalue decomposition of $\mathbf{A}$. But it can be if $\mathbf{U} = \mathbf{V}$, since then the equation on the top is basicaly the eigenvalue decomposition. In that case, we have $\mathbf{AA^T} = \mathbf{A^TA} = \mathbf{\Sigma^2}$. This is a neccesary and sufficient condition for $\mathbf{A}$ to be diagonalisable, or "normal" . In the other case, we say that $\mathbf{A}$ is "deficient".

The summation form in the topmost equation can be used to approximate the matrix $\mathbf{A}$ as a sum of lower rank matrices. this is used in image compression. It's also the value of the spectral norm of the matrix, which is the maximum scaling a vector receives when passed through the matrix. Basically, the spectral norm is :

$$||\mathbf{A}||_2 = \max_{\mathbf{x}}\left(\frac{||\mathbf{Ax}||_\mathbf{2}}{||\mathbf{x}||_\mathbf{2}}\right) = \max_i \sigma_i$$

The last equality is can be proved like this :
$$||\mathbf{Ax}||^\mathbf{2} = \mathbf{x^TA^TAx} = \sum_\mathbf{i} \mathbf{x^T}(\mathbf{u_i}\sigma_\mathbf{i}^\mathbf{2}\mathbf{u_i^T})\mathbf{x} = \sum_\mathbf{i} \sigma_\mathbf{i}^\mathbf{2}(\mathbf{u_i^Tx})^\mathbf{2} \leq \sigma_{\max}^\mathbf{2}\sum(\mathbf{u_i^Tx})^\mathbf{2} = \sigma_{\max}^\mathbf{2}||\mathbf{x}||^\mathbf{2}$$
The last equality is because $\mathbf{U}$ is orthogonal

Eckart-Young theorem

Consider a rank-r matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with the singular value decomposition $\mathbf{A} = \mathbf{U\Sigma V^T} = \sum_{i=1}^{r} \sigma_i\mathbf{u_i v_i^T}$ and the rank-k approximation . For any rank-k matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$ with $k \leq r,$ we have $||\mathbf{A} - \mathbf{B}|| \geq \sigma$ where equality happens when $\mathbf{B} = \sum^k \sigma \mathbf{u v^T}$

nave $||\mathbf{A} - \mathbf{B}||_2 \geq \sigma_{k+1}$, wnere equality nappens wnen $\mathbf{B} = \sum_{i=1} \sigma_i \mathbf{u_i} \mathbf{v_i}^-$.

Proof : Since $\mathbf{B}$ has rank $k < r$, we can construct a vector $\mathbf{w}$ that lives in the null space of $\mathbf{B}$ using the first $(k+1)$ columns of $\mathbf{V}$. Suppose $\mathbf{w} = \alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \cdots + \alpha_{k+1} \mathbf{v_{k+1}} = \mathbf{V_k a}$ where $\mathbf{V_{k+1}}$ is the matrix with $\mathbf{v_i}$ as column vectors and $\mathbf{a}$ is the column vector with entries α_i . Also, suppose that the k linearly independent column vectors as stacked up as row vectors to form $\mathbf{Y}^T$, then the matrix $\mathbf{Y^T V_k}$ has shape $k \times (k+1)$ and thus the equation $\mathbf{Y^T V_k a = 0}$ is solvable (because you can just do Gaussian elimination with some columns not being 0. Then find the specific solution by setting the coordinates in $\mathbf{a}$ to 0 for such columns). Thus there is a vector $\mathbf{w} = \mathbf{V_k a}$ that is orthogonal to all the k linearly independent vectors in $\mathbf{B}$ and is thus in the null space of $\mathbf{B}$.

$$||\mathbf{A} - \mathbf{B}||^2 ||\mathbf{w}||^2 \geq ||(\mathbf{A} - \mathbf{B})\mathbf{w}||^2 = ||\mathbf{Aw}||^2 = \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \cdots + \alpha_{k+1}^2 \sigma_{k+1}^2 \geq \sigma_1^2$$

The first inequality is due to the definition of spectral.

Mind Map

Yes, yes.. this isn't my style, and neither is it yours, but since it's given to us just like that, we might as well ..

Defective

eigenvectors

Non-defective
(diagonalizable)

$A^\top A = AA^\top$ $A^\top A \neq AA^\top$

Normal

Non-normal

$A^\top A = AA^\top = I$

Symmetric
eigenvalues $\in \mathbb{R}$

∃ Inverse Matrix
Regular
(invertible)

Diagonal

Columns are
orthogonal
eigenvectors

Identity
matrix

Positive definite
Cholesky
eigenvalues > 0

Rotation

Orthogonal