Proof of diagonality

Desh Raj

November 18, 2018

Theorem

For an arbitrary matrix D, a diagonal matrix Λ, and a positive scalar γ, the quantity

$$L(D) = \gamma \log |D| + \mathrm{tr}(D\Lambda) - \frac{1}{2}\mathrm{tr}(D^T D)$$

is maximized when D is diagonal.

Lemma 1

If A is an $n \times n$ matrix, then the sum of the n eigenvalues is the trace of A and their product is the determinant of A.

Proof. Let

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Let the eigenvalues be $\lambda_1, \dots, \lambda_n$ and the characteristic polynomial be $p(\lambda) = |\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0 = (\lambda - \lambda_1)\ldots(\lambda - \lambda_n)$, since the eigenvalues are the roots of the characteristic polynomial.

Put $\lambda = 0$ in the characteristic polynomial to get

$$\begin{aligned} &|0I - A| = (0 - \lambda_1)\ldots(0 - \lambda_n) \\ \implies &(-1)^n|A| = (-1)^n\lambda_1\ldots\lambda_n \\ \implies &|A| = \lambda_1\ldots\lambda_n. \end{aligned}$$

Second, consider the coefficient c_{n-1} in the characteristic polynomial. It can be calculated in 2 ways: first, by expanding $p(\lambda) = (\lambda - \lambda_1)\ldots(\lambda - \lambda_n) = -\lambda_1\lambda^{n-1} - \ldots - \lambda_n\lambda^{n-1}$, and so $c_{n-1} = -(\lambda_1 + \ldots + \lambda_n)$.

The coefficient can also be obtained by expanding

$$|\lambda I - A| = \begin{pmatrix} \lambda - a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \lambda - a_{22} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & \lambda - a_{nn} \end{pmatrix}$$

To compute the determinant, we multiply the elements in positions $1j_1, 2j_2, \ldots, nj_n$ for each possible permutation $j_1, \ldots, j_n$ of $1, \ldots, n$. If the permutation is odd, we multiply it by -1 as well. In this way we get $n!$ products, and we add all of them to get the determinant. Of these $n!$ products, one of them is the product $(\lambda - a_n n) \ldots (\lambda - a_{nn})$. All other product terms can contain at most $n - 2$ diagonal elements, which means their degree is at most $n - 2$ in λ. Therefore, the coefficient of λ^{n-1} in the overall determinant comes only from $(\lambda - a_n n) \ldots (\lambda - a_{nn})$, and as done previously, it can be shown that this coefficient is $c_{n-1} = -(a_{11} + \ldots + a_{nn})$. Hence, the sum of eigenvalues is equal to the trace of A. □

Lemma 2

The term tr(AB) under fixed spectrum of A is maximized when the matrices A and B share the same eigenbasis.

Proof. By fixed spectrum, we mean that the eigenvalues of A are fixed. Let $A = U_A \Sigma_A V_A^T$ and $B = U_B \Sigma_B V_B^T$ be the singular value decompositions of A and B. Σ_A is fixed since the spectrum is fixed, and we have to determine the eigenbasis of A.

$$
\begin{aligned}
\mathrm{tr}(AB) &= \mathrm{tr}(BA) \\
&= \mathrm{tr}(U_A \Sigma_A V_A^T U_B \Sigma_B V_B^T) && \text{(invariance under change of basis)} \\
&= \mathrm{tr}(\Sigma_A V_A^T U_B \Sigma_B) && \text{(same argument as above)} \\
&= \mathrm{tr}(\Sigma_B \Sigma_A V_A^T U_B) && \text{(invariance under cyclic permutation)} \\
&= \mathrm{tr}(\Sigma_{AB} V_A^T U_B) && \text{(define } \Sigma_{AB} = \Sigma_B \Sigma_A\text{)}
\end{aligned}
$$

In this expression above, Σ_{AB} is fixed. Furthermore, in the product $V_A^T U_B$, the row i contains the entries $v_i u_1, v_i u_2, \ldots, v_i u_n$. Since the trace only cares about diagonal entries, we can maximize it by setting $v_i = u_i$ for all i. Thus, we get that A should have the same eigenbasis as B. □

Proof of theorem

Proof. Let us fix the spectrum (set of eigenvalues) of D. Using Lemma 1, the first and third terms of $L(D)$ get fixed, and we are only left with the term $\mathrm{tr}(D\Lambda)$. Using Lemma 2, we can see that this term gets maximized when D and Λ share the same eigenbasis. Since Λ is a diagonal matrix, its eigenbasis is the identity matrix I. Therefore, D's eigenbasis must also be I, which implies that D must be a diagonal matrix.

Since the choice of fixed eigenvalues is arbitrary, we can conclude, by peeling argument, that this holds for any D in general. □