Bias-Variance Tradeoff in Low-Rank Representations Proof: Frobenius Norm is the Sum of Squared Singular Values

This proof makes use of the matrix trace operation for square matrices. The trace of a matrix is the sum of the diagonal elements of the matrix. Let $\boldsymbol{B}$ be an M-by-M matrix with elements $B_{i,j}$. Then trace$\{\boldsymbol{B}\} = \text{tr}\{\boldsymbol{B}\} = \sum_{i=1}^{M} B_{i,i}$. One very useful property of the trace operation is that it is invariant to the order of a product of matrices - as long as the products are conformable. Let $\boldsymbol{C}$ be M-by-N with elements $C_{i,j}$ and $\boldsymbol{D}$ be N-by-M with elements $D_{i,j}$. Then both $\boldsymbol{CD}$ and $\boldsymbol{DC}$ are defined. We have

$$\text{tr}\{\boldsymbol{CD}\} = \text{tr}\{\boldsymbol{DC}\}$$

This property follows from the definition of matrix multiplication. The i,i element of $\boldsymbol{CD}$ is $[\boldsymbol{CD}]_{i,i} = \sum_{j=1}^{N} \boldsymbol{C}_{i,j}\boldsymbol{D}_{j,i}$ so

$$\text{tr}\{\boldsymbol{CD}\} = \sum_{i=1}^{M}\sum_{j=1}^{N} \boldsymbol{C}_{i,j}\boldsymbol{D}_{j,i}$$

Similary, the k,k element of $\boldsymbol{DC}$ is $[\boldsymbol{DC}]_{k,k} = \sum_{m=1}^{M} \boldsymbol{D}_{k,m}\boldsymbol{C}_{m,k}$ so

$$\text{tr}\{\boldsymbol{DC}\} = \sum_{k=1}^{N}\sum_{m=1}^{M} \boldsymbol{D}_{k,m}\boldsymbol{C}_{m,k}$$

Interchanging the order of the sums and order of multiplication of the scalars in the sum reveals the equality $\text{tr}\{\boldsymbol{CD}\} = \text{tr}\{\boldsymbol{DC}\}$.

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_M \end{bmatrix}$ be an N-by-M matrix with columns $\boldsymbol{a}_i$. Let the $\boldsymbol{A}$ have singular value decomposition $\boldsymbol{U\Sigma V}^T$ where $\boldsymbol{U}, \boldsymbol{V}$ are square matrices, that is, the full singular value decomposition.

Theorem:

$$||\boldsymbol{A}||_F^2 = \sum_{i=1}^{\min\{M,N\}} \sigma_i^2$$

Proof: First we note that $||\boldsymbol{A}||_F^2 = \sum_{i=1}^{M} \boldsymbol{a}_i^T\boldsymbol{a}_i$ since $\boldsymbol{a}_i^T\boldsymbol{a}_i$ is the sum of the squares of all elements in the i^{th} column of $\boldsymbol{A}$ and we are summing over all M columns.

Next, note that $||\boldsymbol{A}||_F^2 = \sum_{i=1}^{M} \boldsymbol{a}_i^T\boldsymbol{a}_i = \text{tr}\{\boldsymbol{A}^T\boldsymbol{A}\}$ since $\boldsymbol{a}_i^T\boldsymbol{a}_i$ is the i^{th} entry on the diagonal of $\boldsymbol{A}^T\boldsymbol{A}$.

Now substitute the singular value decomposition of $\boldsymbol{A}$ to write

$$||\boldsymbol{A}||_F^2 = \text{tr}\{\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{1}$$
$$= \text{tr}\{\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\} \tag{2}$$

using the orthonormality of the left singular vectors in $\boldsymbol{U}$. Now move $\boldsymbol{V}$ from the left to the right using the fact that the trace is invariant to the order of a product to obtain

$$||\boldsymbol{A}||_F^2 = \text{tr}\{\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\} \tag{3}$$
$$= \text{tr}\{\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\} \tag{4}$$
$$= \sum_{i=1}^{\min\{M,N\}} \sigma_i^2 \tag{5}$$

where the second line follows from the orthonormality of right singular vectors in $\boldsymbol{V}$ and the last line is a consequence of $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$ being a square matrix with the squares of the $\min\{M, N\}$ singular values on the diagonal.