Assignement Deep Learning

hugo.todeschini

January 2019

1 Question 2

We want to prove that $W^* = argmin \|WX - Y\|_F = UV^T$ with W orthgonal matrix and $SVD(YX^T) = U\Sigma V^T$
$\|WX - Y\|_F = \sqrt{tr((WX - Y)(WX - Y)^T)}$
$\|WX - Y\|_F = \sqrt{tr(WX(WX)^T) + tr(YY^T) - tr(Y(WX)^T) - tr(WXY^T))}$
$tr(WX(WX)^T) = \|WX\|_F^2 \, and \, \|WX\|_F = \|X\|_F$ because W is an orthogonal matrix which preserve norms

Moroever $tr(Y(WX)^T) = tr(WXY^T)$ because $tr(A^T) = tr(A)$ for all A matrix.
So $\|WX - Y\|_F = \sqrt{\|X\|_F^2 + \|Y\|_F^2 - 2tr(WXY^T))}$

To minimize $\|WX - Y\|_F$ we need to maximize $tr(WXY^T)$

$SVD(YX^T) = U\Sigma V^T$ So $SVD(XY^T) = V\Sigma U^T$

$tr(WXY^T) = tr(WV\Sigma U^T) = tr(U^T WV\Sigma)$
U,W and V are orthogonal matrix so we can define an orthogonal matrix Z equal to $U^T WV$ and $tr(U^T WV\Sigma) = tr(Z\Sigma)$

Σ is a diagonal matrix with $\Sigma = diag(\sigma_1, \sigma_2, \sigma_3 ...)$ and all $\sigma_i \geq 0$
So tr(ZΣ) = $\sum_{i=1}^{n} \sigma_i z_{i,i}$ and to maximize tr(ZΣ) we have to maximize all $z_{i,i}$ but Z is an orthogonal matrix and all its columns are normed so $z_{i,i} \leq 1$ and if we take all $z_{i,i} = 1$ we get the identity matrix because all $z_{i,j}$ with $i \neq j$ are equal to 0 to have an orthogonal matrix.

So $Z = U^T WV = I$ and $W = UV^T$