Fundamentals of Machine Learning
Exercise 7
Bauer, Steffen & Bhashyam, Anirudh & Löhr, Jessica

February 8, 2021

Question 1

Proof. We begin by proving the lemma. The SVD of X is given by $U\Lambda V$ where U and V are unitary matrices and Λ is a diagonal matrix. Since the matrices are unitary and real, we have, $U^* = U^T = U^{-1}$ and $V^* = V^T = V^{-1}$. Also, $V^T U = UV^T = \mathbf{O}$ (matrices are orthogonal).

$$
\begin{aligned}
(X^T X + \tau \mathbb{I}_D)^{-1} X^T &= (V\Lambda U^T U \Lambda V^T + \tau \mathbb{I}_D)^{-1} V \Lambda U^T \\
&= (V\Lambda^2 V^T + \tau \mathbb{I}_D)^{-1} (U\Lambda^{-1} V^T)^{-1} \\
&= ((U\Lambda^{-1} V^T)(V\Lambda^2 V^T + \tau \mathbb{I}_D))^{-1} \\
&= (U\Lambda V^T + U\Lambda^{-1} V^T \tau \mathbb{I}_D)^{-1} \\
&= V(\Lambda + \Lambda^{-1} \tau \mathbb{I}_D)^{-1} U^T
\end{aligned}
$$

and

$$
\begin{aligned}
X^T (XX^T + \tau \mathbb{I}_N)^{-1} &= V\Lambda U^T (U\Lambda V^T V \Lambda U^T +)^{-1} \\
&= (U\Lambda^{-1} V^T)^{-1} (U\Lambda^2 U^T + t\mathbb{I}_N)^{-1} \\
&= ((U\Lambda^2 U^T + t\mathbb{I}_N)(U\Lambda^{-1} V^T))^{-1} \\
&= (U\Lambda V^T + \tau \mathbb{I}_N U \Lambda^{-1} V^T)^{-1} \\
&= V(\Lambda + \tau \mathbb{I}_N \Lambda^{-1})^{-1} U^T
\end{aligned}
$$

The statement now follows. Using the lemma we can conclude that

$$
\begin{aligned}
\hat{\beta} &= (X^T X + \tau \mathbb{I}_D)^{-1} X^T \mathbf{y} \\
&= X^T (XX^T + \tau \mathbb{I}_N)^{-1} \mathbf{y} \\
&= X^T \hat{\alpha}
\end{aligned}
$$

□

Comment:
Overall well done, the solution is following the suggested approach using SVD and is correctly arriving at the right conclusion. However, the given indentify V^T U = U V^T = 0 is not correct, since the matrices U,V being unitary only implies that V*V = VV* = U*U = UU* = I, where I is the indentity matrix. In fact, if one follows the formulation of SVD from Wikipedia, a decomposition of matrix X with shape (m,n) yields matrices U of shape (m,m), \Lambda of shape (m,n) and V of shape (n,n), which implies that a matrix multiplication of U with V would not even be defined if m!=n, given the specifics of how the SVD is performed. Anyhow, this incorrect indentity is actually not used in the proof, and as such it does not give any problems.