Machine Learning Homework 01

Qiu Yihang

Mar.27, 2022

Proof. We use $\boldsymbol{I}_p$ to denote $p-$dimensional unit matrix, i.e. $\boldsymbol{I}_p = \texttt{diag}(1, 1, ..., 1)_{p\times p}$.

Since $\boldsymbol{V}$ is a $p \times p$ orthogonal matrix, we know $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_p$.

Thus, we have

$$\begin{aligned}
\left(\boldsymbol{Z}^T\boldsymbol{Z} + \lambda\boldsymbol{I}_p\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} &= \left(\left(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T\right)^T\left(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T\right) + \lambda\boldsymbol{V}^T\boldsymbol{V}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} \\
&= \left(\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T + \lambda\boldsymbol{V}^T\boldsymbol{I}_p\boldsymbol{V}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} \\
&= \left(\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{V}^T + \boldsymbol{V}^T\left(\lambda\boldsymbol{I}_p\right)\boldsymbol{V}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} \\
&= \left(\boldsymbol{V}\left(\boldsymbol{D}^2 + \lambda\boldsymbol{I}_p\right)\boldsymbol{V}^T\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y}
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
&\left(\boldsymbol{V}\left(\boldsymbol{D}^2 + \lambda\boldsymbol{I}_p\right)\boldsymbol{V}^T\right)\left(\boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T\right) \\
= \ &\boldsymbol{V}\underset{j}{\texttt{diag}}\left(d_j^2+\lambda\right)\left(\boldsymbol{V}^T\boldsymbol{V}\right)\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T \\
= \ &\boldsymbol{V}\underset{j}{\texttt{diag}}\left(d_j^2+\lambda\right)\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T \ = \ \boldsymbol{V}\boldsymbol{I}_p\boldsymbol{V}^T = \boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_p,
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left(\boldsymbol{Z}^T\boldsymbol{Z} + \lambda\boldsymbol{I}_p\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} &= \left(\boldsymbol{V}\left(\boldsymbol{D}^2 + \lambda\boldsymbol{I}_p\right)\boldsymbol{V}^T\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} \\
&= \boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T\boldsymbol{Z}^T\boldsymbol{y} \\
&= \boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T\left(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T\right)^T\boldsymbol{y} \\
&= \boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{U}^T\boldsymbol{y} \\
&= \boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{1}{d_j^2+\lambda}\right)\underset{i}{\texttt{diag}}\left(d_i\right)\boldsymbol{U}^T\boldsymbol{y} \\
&= \boldsymbol{V}\underset{j}{\texttt{diag}}\left(\frac{d_j}{d_j^2+\lambda}\right)\boldsymbol{U}^T\boldsymbol{y}
\end{aligned}$$

Qed. ∎