Assignment 2 ReadMe

Daniel Bok
ESD
1001049
daniel_bok@mymail.sutd.edu.sg

Wong Yan Yee
ISTD
1001212
yanyee_wong@mymail.sutd.edu.sg

Clement Tan
ESD
1000948
clement_tan@mymail.sutd.edu.sg

January 31, 2017

1 Part 3 Derivation

1.1 Without Regularization

$$
\begin{aligned}
&\arg\min_b \quad \epsilon^2 \\
\epsilon^2 &= \|Ab - c\|^2 \\
&= (Ab - c)^T (Ab - c) \\
&= (b^T A^T - c^T)(Ab - c) \\
&= b^T A^T Ab - b^T A^T c - c^T Ab + c^T c \\
&= b^T A^T Ab - 2c^T Ab + c^T c \qquad (1)
\end{aligned}
$$

We have $b^T A^T c = c^T Ab$ because both of them are identical scalar variables. Differentiating Equation 1, we have

$$
\begin{aligned}
\frac{\partial \epsilon^T \epsilon}{\partial b} &= \frac{\partial b^T A^T A b - 2c^T A b + c^T c}{\partial b} \\
&= -2A^T c + 2A^T A b \qquad (2) \\
&\text{Set Equation 2 to 0} \\
0 &= 2A^T A b - 2A^T c \\
b &= (A^T A)^{-1} A^T c \qquad (3)
\end{aligned}
$$

Notes

By definition, $\frac{\partial b^T V b}{\partial b} = (V + V^T)b$. But since $A^T A$ is symmetric, Equation 2 reduces the term to $2A^T A b$.

1.2 With Regularization

With the regularization term, $\lambda \|b\|_2^2$, the minimization problem becomes

$$
\begin{aligned}
&\arg\min_b \quad \epsilon^2 \\
\epsilon^2 &= \|Ab - c\|^2 + \lambda \|b\|^2 \\
&= (Ab - c)^T (Ab - c) + \lambda b^T b \\
&= (b^T A^T - c^T)(Ab - c) + \lambda b^T b \\
&= b^T A^T A b - b^T A^T c - c^T A b + c^T c + \lambda b^T b \\
&= b^T A^T A b - 2c^T A b + c^T c + \lambda b^T b \qquad (4)
\end{aligned}
$$

Differentiating Equation 4, we have

$$
\begin{aligned}
\frac{\partial \epsilon^T \epsilon}{\partial b} &= \frac{\partial b^T A^T A b - 2c^T A b + c^T c + \lambda b^T b}{\partial b} \\
&= -2A^T c + 2A^T A b + \lambda b \qquad (5) \\
&\text{Set Equation 5 to 0} \\
0 &= 2A^T A b - 2A^T c + \lambda b \\
b &= (A^T A + \lambda I)^{-1} A^T c \qquad (6)
\end{aligned}
$$

Notes

While in Equation 6 we should have taken the term to be $0.5\lambda I$ to be precise, we have chosen not to do so as λ is a constant hyper-parameter. We can thus allow 0.5 to be absorbed into the hyper-parameter with no change to the consistency of Equation 6.