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

February 6, 2017

1 Running the code

We used Python 3.5 to for the assignment. Please also download the package **numpy**. Alternatively, use the environment by Anaconda which has most stuff pre-installed.

Run the file **A2.py** to get the output.

2 Part 3 Derivation

2.1 Without Regularization

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

We have $b^TA^Tc = c^TAb$ because both of them are identical scalar variables. Differentiating Equation 1, we have

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

Notes

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

2.2 With Regularization

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

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

Differentiating Equation 4, we have

$$
\begin{aligned}
\frac{\partial \epsilon^T \epsilon}{\partial b} &= \frac{\partial b^T A^T Ab - 2c^T Ab + c^T c + \lambda b^T b}{\partial b} \\
&= -2A^T c + 2A^T Ab + \lambda b
\end{aligned} \tag{5}
$$

Set Equation 5 to 0

$$
\begin{aligned}
0 &= 2A^T Ab - 2A^T c + \lambda b \\
b &= (A^T A + \lambda I)^{-1} A^T c
\end{aligned} \tag{6}
$$

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.