Machine Learning - Assignment 3

Sudip Sinha

July 18, 2014

Exercise 2

Given

- $VV^{\intercal} = V^{\intercal}V = I$
- D is an invertible diagonal matrix

Proof

Let $B = VD^{-1}V^{\intercal}$

- $AB = VDV^{\intercal}VD^{-1}V^{\intercal} = VD\left(V^{\intercal}V\right)D^{-1}V^{\intercal} = VDID^{-1}V^{\intercal} = V\left(DD^{-1}\right)V^{\intercal} = VIV^{\intercal} = I$
- $BA = VD^{-1}V^{\intercal}VDV^{\intercal} = VD^{-1}\left(V^{\intercal}V\right)DV^{\intercal} = VD^{-1}IDV^{\intercal} = V\left(D^{-1}D\right)V^{\intercal} = VIV^{\intercal} = I$

Thus, $A^{-1} = B = VD^{-1}V^{\intercal}$.

Exercise 3

Let $f(w) = \|X\boldsymbol{w} - \boldsymbol{y}\|^2$

We just have to show that the square of the norm is a convex function.

$$\begin{aligned}
\|(1-t)x + ty\|^2 &= (1-t)^2\|x\|^2 + t^2\|y\|^2 + 2t(1-t)\langle x, y\rangle \qquad \forall t \in [0,1] \\
&= (1-t)\|x\|^2 + t\|y\|^2 - t(1-t)\left(\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle\right) \\
&= (1-t)\|x\|^2 + t\|y\|^2 - t(1-t)\|x-y\|^2 \\
&\leq (1-t)\|x\|^2 + t\|y\|^2 \qquad \text{Since } t(1-t) \geq 0 \text{and } \|\cdot\|^2 \geq 0
\end{aligned}$$

Thus, the least squares loss function $\|y - Xw\|^2$ is a convex function of w.

Exercise 4

Let $W = \operatorname{diag}(r_i)$. Therefore, $W^{\intercal} = W$

Thus, the problem can be reformulated as

$$\begin{aligned}
w &= \arg\min_{w}(e) \\
\text{where } e &= (y - Xw)^{\intercal}W(y - Xw) \\
&= \left(y^{\intercal} - w^{\intercal}X^{\intercal}\right)W(y - Xw) \\
&= y^{\intercal}Wy - y^{\intercal}WXw - w^{\intercal}X^{\intercal}Wy + w^{\intercal}X^{\intercal}WXw
\end{aligned}$$

Using the rules of Matrix Calculus given here:

$$\begin{aligned}
\frac{\partial e}{\partial w} &= -\left(y^{\intercal}WX\right)^{\intercal} - X^{\intercal}Wy + \left(X + X^{\intercal}\right)w \\
&= \left(X^{\intercal}WX + \left(X^{\intercal}WX\right)^{\intercal}\right)w - X^{\intercal}W^{\intercal}y - X^{\intercal}Wy \\
&= 2\left(X^{\intercal}WXw - X^{\intercal}Wy\right) \qquad (\text{Since } W^{\intercal} = W) \\
&= 2X^{\intercal}W(Xw - y)
\end{aligned}$$

For minimum error:

$$\begin{aligned} \frac{\partial e}{\partial w} &= 0 \\ \implies X^T W X w &= X^T W y \\ \implies w &= \left(X^T W X\right)^{+} X^T W y \end{aligned}$$

Exercise 6

Notations

Symbol	Dimension	Description
X	$m \times n$	The design matrix
y	$m \times 1$	The actual values of the predicted variable
$O(\cdot)$	-	The Big O notation

Computational Complexity[1]

LLS

Output	Complexity	Operation
$X^T X$	$O\left(mn^2\right)$	Matrix multiplication
$X^T y$	$O(mn)$	Matrix multiplication
$X^T X w = X^T y$	$O\left(n^3\right)$	Solving linear system of equations

Now, we shall assume that $n \ll m$. In this case, $O\left(mn^2\right)$ dominates over $O\left(n^3\right)$. Thus the computational complexity of LLS is $O\left(mn^2\right)$.

kNN

Output	Complexity	Operation
$D_i = \Sigma_i \left(X_{test} - X_{train;i}\right)^2$	$O(mn)$	Subtraction, squaring and summing (each $O(n)$) / row of train
Sorted index vector δ	$O(m \ln m)$	Sorting
Predicted class	$O(k)$	Mode

Now, we shall assume that $k \ll m$ and $\ln m \ll n$. In this case, $O(mn)$ dominates over the other complexities.

Space Complexity

LLS

Object	Space
X	mn
y	m
X^T	mn
$X^T X$	n^2
$X^T y$	n
w	n

Thus the space complexity of LLS is $O(mn)$.

kNN

Object	Space
X	mn
y	m
$X^T y$	n
δ	n

Thus the space complexity of LLS is $O(mn)$.

Final results

	Computation	Time
LLS	$O\left(mn^2\right)$	$O(mn)$
kNN	$O(mn)$	$O(mn)$

[1]Taken from Computational complexity of mathematical operation in Wikipedia