Machine Learning - Assignment 3

Sudip Sinha

May 4, 2014

Exercise 2

Given

- $VV^T = V^T V = I$
- D is an invertible diagonal matrix

Proof

Let $B = VD^{-1}V^T$

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

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

Exercise 3

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

By the definition of norm,

$$\begin{aligned} \|tx + (1-t)\,y\| &\leq \|tx\| + \|(1-t)\,y\| = t\,\|x\| + (1-t)\,\|y\| \\ \implies \|tx + (1-t)\,y\|^2 &\leq (t\,\|x\| + (1-t)\,\|y\|)^2 \leq t^2\,\|x\|^2 + (1-t)^2\,\|y\|^2 \end{aligned}$$

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

Exercise 4

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

Thus, the problem can be reformulated as

$$w = \arg\min_w (e)\,,\ \text{where } e = (y - Xw)^T\,W\,(y - Xw)$$

Using the rules of Matrix Calculus given here:

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

For minimum error:

$$\begin{aligned} & \tfrac{\partial e}{\partial w} && = 0^T \\ \implies & 2\left(X^TW\,(Xw - y)\right)^T && = 0^T \\ \implies & X^TW\,(Xw - y) && = 0 \\ \implies & X^TWXw && = X^Ty \\ \implies & w && = \left(X^TWX\right)^{-1}X^Ty \end{aligned}$$

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^TX	$O\left(mn^2\right)$	Matrix multiplication
X^Ty	$O(mn)$	Matrix multiplication
$X^TXw = X^Ty$	$O\left(n^3\right)$	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^TX	n^2
X^Ty	n
w	n

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

kNN

Object	Space
X	mn
y	m
X^Ty	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 operations in Wikipedia