CSE 404 HW4

Sayem Lincoln

A54207835

1) The linear regression can be expressed as to the

$$E(w) = \frac{1}{N} ||Xw - y||^2$$

Minimizing the function with respect to w leads to the optimal w as $(x^T x)^{-1} x^T y$. This solution holds only when $x^T x$ is non singular. The objective function is commonly minimized instead

$$E_2(w) = ||Xw - y||^2 + \lambda ||w||^2$$

where $\lambda > 0$

$$E_2(w) = (xw - y)(xw - y) + \lambda(ww)$$

$$= wxxw - 2wxy - yy + \lambda(w'w)$$

Differentiating w to minimize $E_2(w)$, we get

$$2xxw - 2xy + \lambda w = 0$$

$$\Rightarrow (2xx + \lambda) w = 2xy$$

$$\Rightarrow w = inv(2xx + \lambda) + 2xy$$

Earlier the problem was that xx could be singular. As $\lambda > 0$ and user defined we can use λ such that $2xx + \lambda$ is not singular and hence invertible. Hence a solution to w can be found as $inv(x'x + \lambda)$ exists.

2) The hat matrix $H = X(X^T X)^{-1} X^T$, given $X^T X$ is invertible.

a) H is symmetric if $H = H^T$ we have

$$H = X(X^T X)^{-1} X^T$$

$$H^T = [X(X^T X)^{-1} \cdot X^T]^T$$

$$= (X^T)^T ((X^T \cdot X)^{-1} X)^T$$

$$= X(X^T \cdot X)^{-1} X^T \quad (\text{Here } X^T X \text{ is invertible})$$

$H^T = H$ — hence H is symmetric.

b) Show that $H^k = H$, for any positive integer k.

Let $k = 2$

H is a nxn square matrix and it is independent which can be verified as follows $\Rightarrow H^2 = H \cdot H$

$$HH = (X(X^T X)^{-1} X^T)(X(X^T X)^{-1} X^T)$$

$$= X(X^T X)^{-1} (X^T X)(X^T X)^{-1} X^T = X(X^T X)^{-1} X^T$$

Here $(X^T X)(X^T X)^{-1} = I$ because $A \cdot A^{-1} = I$)

so, $X(X^T X)^{-1} X^T = H$

Hence for any positive integer k, $H^k = H$. Hence proved.

c) Show that $(I-H)^k = I-H$ (for any positive integer k)

Here I is ony identity matrix of size N.

Similarly, as in part b,

Let $k=2$

$$(I-H)^2 = (I-H)(I-H) = I \cdot I - 2HI + H \cdot H = I - 2H + H = I - H$$

$(H \cdot H = H \rightarrow$ proved in part b)

Hence for any positive integer k, $(I-H)^k = I - H$. Hence proved.

d) Show that trace$(H) = d+1$, where the trace is the sum of diagonal elements.

It is known that trace(AB) = trace(BA) — (1)

Now,

$$\text{trace}(H) = \text{trace}\left(X(X^T \cdot X)^{-1} X^T\right)$$
$$= \text{trace}\left((X^T X)^{-1} X^T \cdot X\right) \text{ — from (1)}$$
$$= \text{trace}(I)$$

because $A^{-1} \cdot A = I$

$(X^T X)^{-1} (X^T \cdot X) = I$

Then trace$(H) = d+1$. Hence proved.