Instituto Politécnico Nacional

Centro de Investigación en cómputo

Chapter 3: The linear model.

Subject: Introducción a machine learning

Student:

- Carpintero Mendoza Marcos Mauricio

Profesor:
Menchaca Mendez Ricardo

14/10/19

Índice

1. Exercises.

1.1. Linear regression.

1.1.1. The algorithm.

Exercise 3.3

Consider the hat matrix $H = X(X^TX)^{-1}X^T$, where X is an B by $d+1$ matrix, and X^TT is invertible.

a) Show that H is symmetric.

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

c) If I is the identity matrix of size N, show that $(I - H)^k = I - H$ for any possible positive integer k.

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

Solution:

a) First, we want to know that $H = H^T$.

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

Using the properties of transposed matrix.

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

b) We want to demonstrate that

$$H^k = \prod_{i=1}^{k} H = H$$

$$\prod_{i=1}^{k} H = \prod_{i=1}^{k} X(X^TX)^{-1}X^T$$

- Let us prove: k = 1

$$\prod_{i=1}^{1} X(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T$$

- Now, our induction hypothesis is based on k = n. Let us assume it is true.

$$\prod_{i=1}^{n} X(X^T X)^{-1} X^T = X(X^T X)^{-1} X^T$$

- Let us prove: k = n+1

$$\prod_{i=1}^{n+1} X(X^T X)^{-1} X^T$$

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

- By our hypothesis:

$$= X(X^T X)^{-1} X^T \cdot X(X^T X)^{-1} X^T$$
$$= (X(X^T X)^{-1} X^T)^2$$

- We know that: $H = H^2$.

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

- Finally:

$$H^k = H$$

c) We want to demonstrate that

$$(I - H)^k = \prod_{i=1}^{k} (I - H) = (I - H)$$

- Let us prove k = 1.

$$\prod_{i=1}^{1} (I - H) = I - H$$

- Now our induction hypothesis is: k = n. Let us assume it true.

$$\prod_{i=1}^{n} (I - H) = I - H$$

- Demonstrate for k = n + 1.

$$\prod_{i=1}^{n+1} (I - H) = I - H$$

$$\prod_{i=1}^{n+1} (I - H) = \prod_{i=1}^{n} (I - H) \cdot (I - H)$$
$$= (I - H) \cdot (I - H) = (I - H)^2$$
$$= I^2 - HI - HI + H^2 = I - H - H + H = I - H$$

- Finally:

$$(I - H)^k = (I - H)$$

c) First: We know that $trace(AB) = trace(BA)$

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

$$= trace(X^TX(X^TX)^{-1}) = trace((X^TX)(X^TX)^{-1})$$

$$= trace(I) = \sum_{n=1}^{d+1} i_{nn} = d + 1$$

1.2. Nonlinear transformation.

1.2.1. The $\mathcal{Z}$ space.

Exercise 3.12

We know that in the Euclidean plane, the perceptron model $\mathcal{H}$ cannot implement all 16 dichotomies on 4 points. That is, $m_{\mathcal{H}}(4) < 16$. Take the feature transform Φ in (3.12).

1. Show that $m_{\mathcal{H}_\Phi}(3) = 8$.
2. Show that $m_{\mathcal{H}_\Phi}(4) < 16$.
3. Show that $m_{\mathcal{H} \cup \mathcal{H}_\Phi}(4) = 16$.

That is, if you used lines, $d_{\text{vc}} = 3$; if you used elipses, $d_{\text{vc}} = 3$; if you used lines and elipses,$d_{\text{vc}} > 3$.

Solution:

The transformation is: $\Phi(x) = (1, x_1^2, x_2^2)$. Now the weights in the space '$\mathcal{Z}$' are $W_{\text{PLA}} = (-0.6, 0.6, 1)$ it leads us to hyphotesis $g(x) = sign(1 - 0.6x_1^2 + 0.6x_2^2)$, finally it became a 2D perceptron.

1. If $m_{\mathcal{H}_\Phi}(3) = 8$, then the breakpoint is: $k = 4$.
   In the previous chapter we found out the formula for convex sets.

$$m_{\mathcal{H}}(N) = 1 + \frac{N^3}{6} + \frac{5N}{6}$$

$$m_{\mathcal{H}_\Phi}(3) = 1 + \frac{27}{6} + \frac{15}{6}$$

$$m_{\mathcal{H}_\Phi}(3) = 8$$

2. Using the following theorem:

$$\text{If } m_h(k) < 2^k \quad \Longrightarrow \quad m_h(N) \leqslant \sum_{i=0}^{k-1} \binom{N}{i}$$

$$m_{\mathcal{H}_\Phi}(4) \leqslant \sum_{i=0}^{3} \binom{N}{i}$$

$$m_{\mathcal{H}_\Phi}(4) \leqslant 15 < 16$$

3. In the previous chapter we found out:

$$m_{\mathcal{H}_\Phi}(4) = 14$$

   The two missing correct classified dichotomies now can be shatter using the elipses, then:

$$m_{\mathcal{H} \cup \mathcal{H}_\Phi}(4) = 14 + 2 \geqslant 16$$

   But we also have

$$m_{\mathcal{H} \cup \mathcal{H}_\Phi}(4) \leqslant 16$$

   Finally:

$$m_{\mathcal{H} \cup \mathcal{H}_\Phi}(4) = 16$$