Inductive Learning

Hanlin He*

September 8, 2017

1

$$
\begin{aligned}
\frac{\partial}{\partial \theta_i} J(\theta_i) &= \frac{1}{m} \sum_{i=1}^{i=m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \\
\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) &= \frac{1}{4} \left((\theta_0 + 3\theta_1 - 2) + (\theta_0 + \theta_1 - 2) + (\theta_0 - 1) + (\theta_0 + 4\theta_1 - 3) \right) \\
&= \theta_0 + 2\theta_1 - 2 \\
\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) &= \frac{1}{4} \left(3(\theta_0 + 3\theta_1 - 2) + (\theta_0 + \theta_1 - 2) + 0(\theta_0 - 1) + 4(\theta_0 + 4\theta_1 - 3) \right) \\
&= 2\theta_0 + 6.5\theta_1 - 5
\end{aligned}
$$

Let θ_0^i denotes the ith round value of θ_0, $\alpha = 0.1$.

- Round 0

$$\theta_0^0 = 0 \tag{1}$$
$$\theta_1^0 = 1 \tag{2}$$

- Round 1

$$\theta_0^1 = 0 - 0.1 \times (0 + 2 - 2) = 0 \tag{3}$$
$$\theta_1^1 = 1 - 0.1 \times (2 \times 0 + 6.5 \times 1 - 5) = 0.85 \tag{4}$$

*hxh160630@utdallas.edu

- Round 2

$$\theta_0^2 = 0 - 0.1 \times (0 + 2 \times 0.85 - 2) = 0.03 \tag{5}$$
$$\theta_1^2 = 0.85 - 0.1 \times (2 \times 0 + 6.5 \times 0.85 - 5) = 0.7975 \tag{6}$$

- Round 3

$$\theta_0^3 = 0.03 - 0.1 \times (0.03 + 2 \times 0.7975 - 2) = 0.0675 \tag{7}$$
$$\theta_1^3 = 0.7975 - 0.1 \times (2 \times 0.03 + 6.5 \times 0.7975 - 5) = 0.773125 \tag{8}$$

- Round 4

$$\theta_0^4 = 0.0675 - 0.1 \times (0.0675 + 2 \times 0.773125 - 2) = 0.106125 \tag{9}$$
$$\theta_1^4 = 0.773125 - 0.1 \times (2 \times 0.0675 + 6.5 \times 0.773124 - 5) = 0.7571 \tag{10}$$

- Round 5

$$\theta_0^5 = 0.106125 - 0.1 \times (0.106125 + 2 \times 0.7571 - 2) = 0.1441 \tag{11}$$
$$\theta_1^5 = 0.7571 - 0.1 \times (2 \times 0.106125 + 6.5 \times 0.7571 - 5) = 0.7438 \tag{12}$$

2

- False Positive: 10%
- False Negative: 20%

3

a) Specific hypothesis (S)

Pros Less false positive error;

Cons More false negative errors.

b) General hypothesis (G)

Pros Less false negative errors.

Cons More false positive error;

4

Definition 1. A hypothesis h is *consistent* with a set of training examples D if and only if $h(x) = c(x)$ for each example $\langle x, c(x) \rangle$ in D.

$$Consistent(h, D) \equiv (\forall \langle x, c(x) \rangle \in D) h(x) = c(x)$$

Definition 2. The *version space*, denoted $VS_{H,D}$, with respect to hypothesis space H and training examples D, is the subset of hypotheses from H consistent with the training examples in D.

$$VS_{H,D} \equiv \{h \in H | Consistent(h, D)\}$$

5

The most general hypothesis has (**?**) value for each attribute.

6

a) $|X| = 3^4 = 81$.

b) $(instance) = 2^{16}$.

c) $\binom{4}{2} \times 4 = 24$

7

$$h_0 \leftarrow \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$$
$$h_1 \leftarrow \langle 1, 1, 0, 1, 1 \rangle$$
$$h_2 \leftarrow \langle 1, 1, 0, 1, 1 \rangle$$
$$h_3 \leftarrow \langle 1, 1, ?, 1, ? \rangle$$
$$h_4 \leftarrow \langle 1, 1, ?, 1, ? \rangle$$
$$h_5 \leftarrow \langle 1, 1, ?, 1, ? \rangle$$

8

$$(GPA < 3.5 \wedge Exp \geq 3) \vee (GPA \geq 3.5 \wedge Exp \geq 1) \tag{13}$$

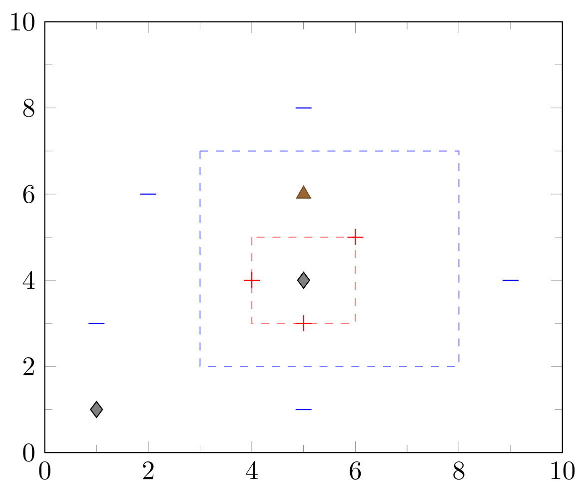
9

a) $S \equiv (4 \leq x \leq 6) \wedge (3 \leq y \leq 5)$, shown as red dash line square in fig. 1. Points on the line should be included.

b) $G \equiv (3 \leq x \leq 8) \wedge (2 \leq y \leq 7)$, shown as blue dash line square in fig. 1. Points on the line should be excluded.

c) A query lying between S and G would guarantee to reduce the size of the version space, for example $x = 5, y = 6$, shown as triangle in fig. 1.
A query outside G or inside S would not reduce size of the version space, for example $x = 1, y = 1$ and $x = 5, y = 4$, shown as diamond in fig. 1.

Figure 1: S and G in Diagram



d) Four training sample would be sufficient to achieve a particular target concept via CANDIDATE-ELIMINATION algorithm.
Take concept $3 \leq x \leq 5, 2 \leq y \leq 9$ as an example. The following four training sample would be sufficient.

+ (3,2)

+ (5,9)

- (2,1)

- (6,10)

10

a) Trace:

i. Original hypothesis.

ii. + $\langle\langle male\ brown\ tall\ US\rangle\langle female\ black\ short\ US\rangle\rangle$

$$S \equiv \langle\langle male\ brown\ tall\ US\rangle\langle female\ black\ short\ US\rangle\rangle \tag{14}$$

$$G \equiv \langle\langle ?\ ?\ ?\ ?\rangle\langle ?\ ?\ ?\ ?\rangle\rangle \tag{15}$$

iii. + $\langle\langle male\ brown\ short\ French\rangle\langle female\ black\ short\ US\rangle\rangle$

$$S \equiv \langle\langle male\ brown\ ?\ ?\rangle\langle female\ black\ short\ US\rangle\rangle \tag{16}$$

$$G \equiv \langle\langle ?\ ?\ ?\ ?\rangle\langle ?\ ?\ ?\ ?\rangle\rangle \tag{17}$$

iv. $-$ $\langle\langle female\ brown\ tall\ German\rangle\langle female\ black\ short\ Indian\rangle\rangle$

$$S \equiv \langle\langle male\ brown\ ?\ ?\rangle\langle female\ black\ short\ US\rangle\rangle \tag{18}$$

$$G \equiv \langle\langle male\ ?\ ?\ ?\rangle\langle ?\ ?\ ?\ ?\rangle\rangle, \langle\langle ?\ ?\ ?\ ?\rangle\langle ?\ ?\ ?\ US\rangle\rangle \tag{19}$$

v. + $\langle\langle male\ brown\ tall\ Irish\rangle\langle female\ brown\ short\ Irish\rangle\rangle$

$$S \equiv \langle\langle male\ brown\ ?\ ?\rangle\langle female\ ?\ short\ ?\rangle\rangle \tag{20}$$

$$G \equiv \langle\langle male\ ?\ ?\ ?\rangle\langle ?\ ?\ ?\ ?\rangle\rangle \tag{21}$$

b) $2^8 = 256$

c) Each time specify a query with one attribute with mark '-' would reduce the general hypothesis by half or some part.

d)