Assignment7 – Learning

Given: June 13 Due: June 18

Problem 7.1 (Loss)

Our goal is to find a linear approximation $h(x) = ax$ for the series of square numbers $0, 1, 4, 9, 16$.

1. Model this situation as an *inductive learning problem*.

Solution: The *inductive learning problem* is $(\mathcal{H}, f)$ where

- the hypothesis space $\mathcal{H}$ is the set containing all functions $h(x) = ax$ with $\mathrm{dom}(h) = \{0, \ldots, 4\}$ for $a \in \mathbb{R}$
- the target function is $f(x) = x^2$ with $\mathrm{dom}(f) = \{0, 1, \ldots, 4\}$

---

2. Assuming all 5 possible examples are equality probable, compute the generalized loss using the *squared error loss* function. (This is a function of h.)

Solution: Each example (x, x^2) has probability $1/5$. For each x, the loss is $L_2(x^2, ax) = (x^2 - ax)^2$. Thus for each $h(x) = ax$, we have

$$GenLoss(h) = \sum_{x=0,\ldots,4} (x^2-ax)^2 \cdot 1/5 = ((1-a)^2+(4-2a)^2+(9-3a)^2+(16-4a)^2)/5 = (354-200a+30a^2)/5$$

---

3. Find h^*.

Solution: We need to find the a that minimizes the loss. The derivative of $GenLoss$ for a is $(60a - 200)/5$. So the minimum is at $a = 10/3$.

---

4. What is the *error rate* of h^*?

Solution: The error rate is $4/5 = 1$ because $h^*(x) = 10x/3$ predicts 4 out of 5 examples incorrectly. (E.g., $h(x) = x$ would have better error rate $3/5$ despite having higher generalized loss.)

Problem 7.2 (Overfitting)

Explain what *overfitting* means and why we want to avoid it.

Solution: Overfitting is a modeling error that occurs when the chosen hypothesis is too closely fit to a sample set of data points. It picks an overly complex hypothesis that also explains idiosyncrasies and errors in the data. A simpler hypothesis that fits the data less exactly is often a better match for the underlying mechanisms.

Problem 7.3 (Decision List)

We want to construct a decision list to classify the data below where result values V depend on 4 attributes A, B, C, D. The tests should be conjunctions of literals.

1. Assume your literals must be of the form $attribute = number$. Which values of k allow for giving the shortest possible decision list in k-DL (i.e., using at most k literals per test)? Give one such list.

   Solution: For $k = 3$ (or higher), we can build a list of length 2:

   if $A = 0 \wedge C = 1 \wedge D = 0$ then 0 else
   if $A = 0 \wedge C = 0 \wedge D = 1$ then 0 else
   1.
   Technically not correct but also interesting is the following solution for $k = 2$ with a list of length 3:
   if $A = 1$ then 1 else
   if $C = 0 \wedge D = 0$ then 1 else
   if $C = 1 \wedge D = 1$ then 1 else
   0

   ---

2. Now assume your literals may also be of the form $attribute = attribute$. Answer the same question as above.

   Solution: For $k = 1$, we can build a list of length 2:

   if $A = 1$ then 1 else
   if $C = D$ then 1 else
   0

Example	A	B	C	D	V
#1	1	0	0	0	1
#2	1	0	1	1	1
#3	0	1	0	0	1
#4	1	1	0	1	1
#5	0	0	1	1	1
#6	0	1	1	0	0
#7	0	1	0	1	0
#8	0	0	1	0	0