Assignment7 – Learning

Given: June 12 Due: June 23

Problem 7.1 (Decision Tree Learning in Python)

Implement the *Decision Tree Learning algorithm* (*DTL*) in Python using the files at `https://kwarc.info/teaching/AI/resources/AI2/dtl`.

Solution: See `https://kwarc.info/teaching/AI/resources/AI2/dtl`

Problem 7.2 (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 $\mathrm{dom}(h) = \{0, \dots, 4\}$ for $a \in \mathbb{R}$
   - the target function is $f(x) = x^2$ with $\mathrm{dom}(f) = \{0, 1, \dots, 4\}$

   ---

2. Assuming all 5 possible examples are equally 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

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

   ---

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 7.3 (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 7.4 (Competition (due September 15))

In this competition, you will implement an agent that explores the FAULumpus world. You will receive up to 2 percentage points of additional bonus for your agents.

All further details will be posted on studon soon.