task_10.1

Machine Learning (WiSe 2025/2026)

Author: Suvansh Shukla
Matriculation No. 256245

Assignment 10 Task 1

Part A

Active learning systems attempt to overcome the labeling bottleneck by asking
queries in the form of unlabeled instances to be labeled by an oracle (e.g., a human
annotator). In this way, the active learner aims to achieve high accuracy using
as few labeled instances as possible, thereby minimizing the cost of obtaining
labeled data[1]

Active learning can be used to build a case base by taking cases the model is unsure of then
asking the *oracle* to label them. Slowly one-by-one a case base can be constructed by adding
more and more of these queried cases.

Part B

These are two examples of measures, that can be used in an active learning context:

Entropy: represents the measure of uncertainty or impurity.

$$x^*_{ENT} = \underset{x}{argmax} - \sum_i P(y_i|x;\theta) log P(y_i|x;\theta)$$

Confidence: represents how sure the model is about it's prediction.

$$x^*_{LC} = \underset{x}{argmin} P(y^*|x;\theta)$$

Where, $y^* = argmax_y P(y|x;\theta)$ is the most likely class labeling.

Part C

The following are some of the differences between such an active learner and
an algorithm like IB2:

- Requires a measure like Entropy or Confidence
- Requires an oracle that can answer queries or provide labels
- Handles noisy data better by selectively querying only the most informative examples for labeling

---

---

1. https://research.cs.wisc.edu/techreports/2009/TR1648.pdf ↩