

Data Analysis Mini-Project

Omer Zilcer- 318247350

Omer Geva - 316551472

Part A

Our Implementation

- implemented teacher as an abstract class, and each Teacher implementation inherits from it.
- Both Teacher and learner received the full database of instances (X) and the teacher also received the distribution to labels (Y). The learner receives a shuffled version of the X table so we can iterate over it in a random order.
- each element in the X dataset is represented as [item_index, binary features numpy array, name].
- We implemented a Conjunction class that saves the item that represents it, numpy arrays of the predicates(offsets, values) and also statistics on hits and misses during runtime.
- the code was implemented using numpy arrays and API for optimal performance (taking advantage of the CPU vector capabilities).
- The decision list is a list of conjunctions that we maintain during the run of the algorithm.

The Teachers

- The difference between the two teachers (TeacherA and TeacherB) is the way each one chooses the discriminative feature in the case of a wrong guess.
- TeacherA simply looks for all the discriminative features between the current instance ('data'), and the example instance of the guessed label ('base_data'), and randomly chooses one of them.

```
class TeacherA(Teacher):  
    def get_new_feature(self, data, conjunction):  
        # find different features between data1 and data2  
        existing_features = conjunction.get_predicate_indices()  
        base_data = conjunction.get_base_data()  
        diff_features = self._get_diff_features(data, base_data, existing_features)  
        assert len(diff_features) > 0  
  
        # choose random feature from the diff  
        idx = rnd.randint(0, len(diff_features) - 1)  
        feature_idx = diff_features[idx]  
        return feature_idx
```

- TeacherB calculates the percentage of occurrences of each feature in each label in its initialization.
- Each time we want to choose a discriminative feature between two instances we choose the one that has the biggest difference in percentages of occurrence in the different labels.

```

class TeacherB(Teacher):
    def __init__(self, X, Y):
        super().__init__(X, Y)
        # build stats for each type
        self.types_stats = {}
        self.types_count = {}
        features_shape = data_get_features(X[0]).shape
        for data in X:
            label = self.get_correct_label(data)
            if not label in self.types_stats:
                self.types_stats[label] = np.zeros(features_shape, dtype=np.int32)
                self.types_count[label] = 0
            self.types_stats[label] = np.add(self.types_stats[label], data_get_features(data))
            self.types_count[label] += 1
        for label in self.types_stats:
            self.types_stats[label] = self.types_stats[label] / self.types_count[label]

    def get_new_feature(self, data, conjunction):
        # find different features between data1 and data2
        existing_features = conjunction.get_predicate_indices()
        base_data = conjunction.get_base_data()
        diff_features = self._get_diff_features(data, base_data, existing_features)
        assert len(diff_features) > 0

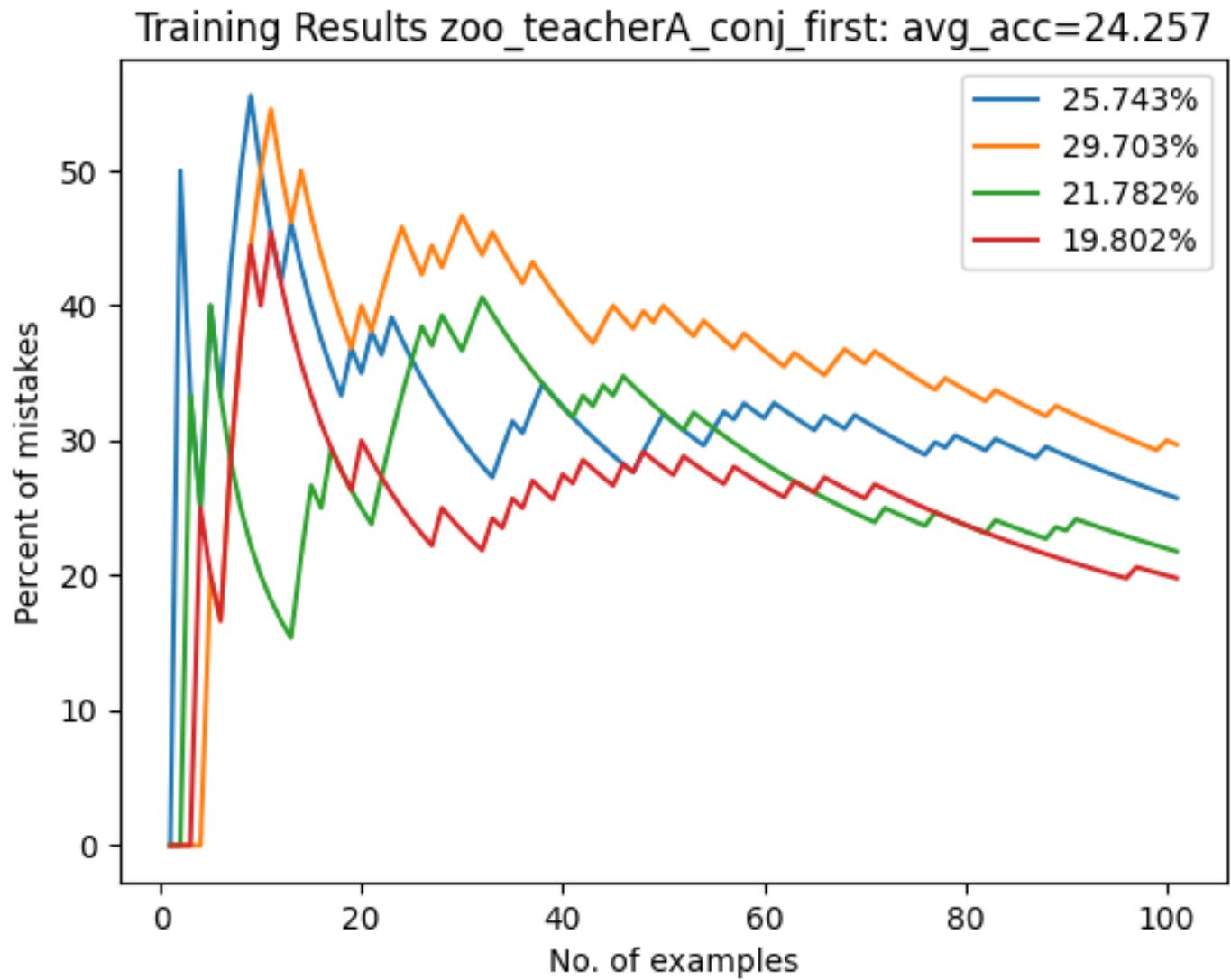
        # find the best feature that has the biggest delta between labels
        data_stats = self.types_stats[self.get_correct_label(data)]
        base_data_stats = self.types_stats[self.get_correct_label(base_data)]
        diff_arr = np.abs(data_stats[diff_features] - base_data_stats[diff_features])
        max_idx = np.argmax(diff_arr)

        return diff_features[max_idx]

```

Part A

- Zoo:
- 4 runs of TeacherA



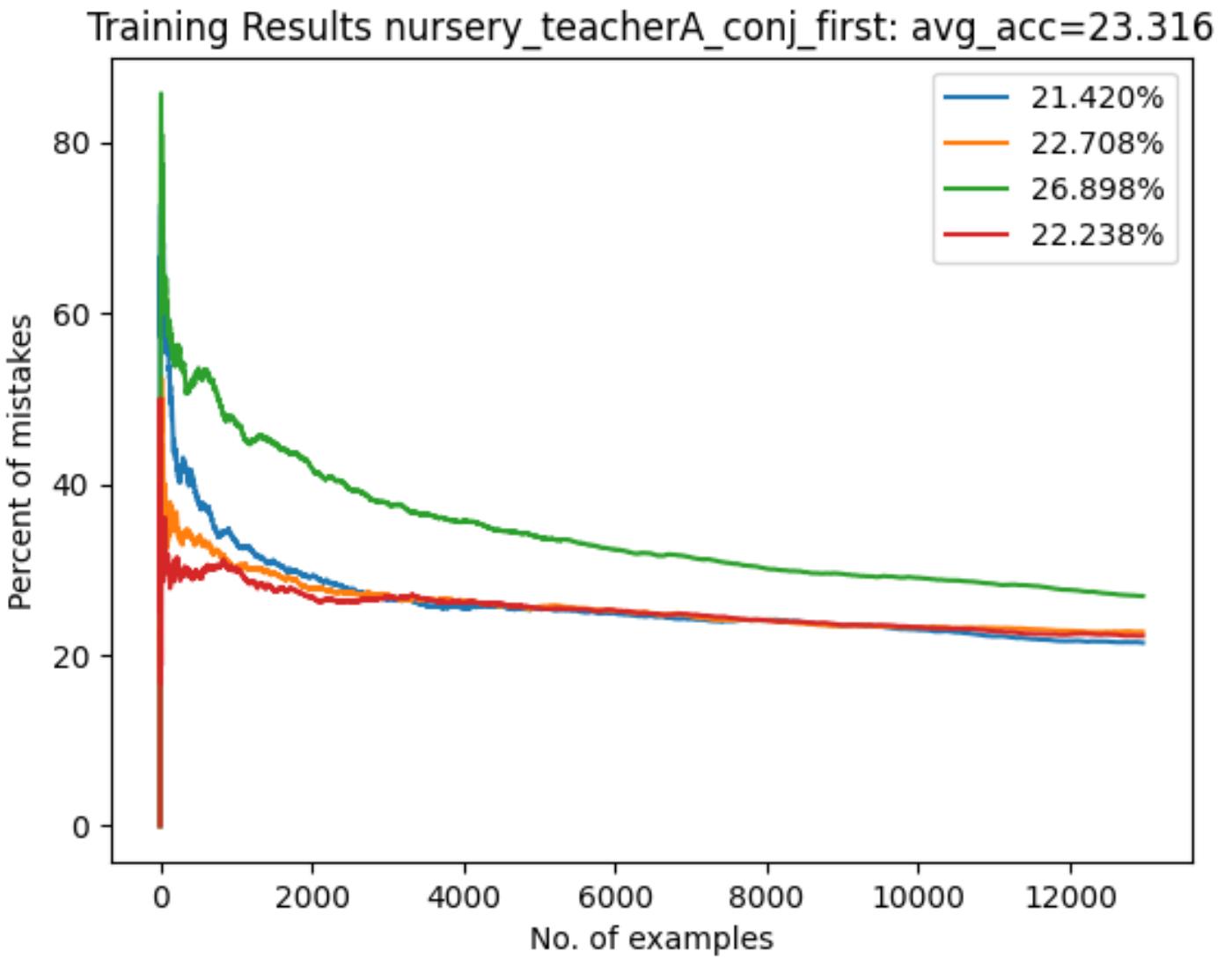
Part A

- Zoo:
- Example of a final decision list TeacherA

```
[(9, 1), (7, 1), (2, 0)] => Label: 1
[(1, 1)] => Label: 2
[(11, 0), (6, 1), (12, 1), (2, 1), (10, 1)] => Label: 3
[(14, 1), (15, 0), (12, 0)] => Label: 7
[(7, 0), (6, 1), (10, 1)] => Label: 7
[(11, 0), (10, 1), (4, 0)] => Label: 3
[(11, 0), (17, 1), (8, 1), (7, 1)] => Label: 5
[(15, 0), (19, 0)] => Label: 7
[(12, 0), (0, 1)] => Label: 6
```

Part A

- Nursery:
- 4 runs of TeacherA



[(), 0, (13, 1), (9, 0), (7, 0), (21, 1), (11, 1), (6, 1), (17, 1), (1, 0), (24, 1)] => Label: not_recom
[(), 25, 0, (9, 0), (1, 0), (14, 0), (7, 0), (22, 0), (26, 1), (0, 1), (3, 0), (23, 1), (4, 0)] => Label: priority
[(), 20, 1, (1, 1), (15, 0), (21, 1), (18, 0), (6, 1), (9, 1), (24, 0), (12, 1), (26, 0)] => Label: priority
[(), 8, 1, (5, 1), (17, 0), (14, 1), (18, 1), (0, 0), (24, 0), (21, 1), (12, 0), (15, 1)] => Label: priority
[(), 18, 0, (22, 0), (24, 0), (10, 0), (9, 1), (15, 1), (3, 0), (16, 1), (20, 0), (2, 1), (26, 0), (5, 1)] => Label: priority
[(), 3, 1, (2, 1), (18, 0), (14, 0), (22, 1), (17, 1), (13, 0), (10, 0), (8, 0), (25, 0), (26, 0)] => Label: not_recom
[(), 13, 0, (6, 1), (20, 1), (22, 1), (15, 1), (10, 0), (1, 1), (24, 0)] => Label: priority
[(), 3, 1, (12, 0), (9, 1), (17, 0), (24, 0), (23, 1), (18, 1), (15, 0), (13, 1), (25, 0)] => Label: priority
[(), 21, 0, (24, 1)] => Label: not_recom
[(), 10, 0, (14, 0), (24, 1)] => Label: not_recom
[(), 13, 0, (22, 1), (18, 1), (10, 1), (14, 0), (0, 0), (19, 1), (2, 1), (3, 0), (7, 0)] => Label: priority
[(), 22, 1, (1, 0), (6, 1), (0, 0)] => Label: priority
[(), 20, 1, (9, 1), (5, 1), (15, 0), (16, 0), (23, 0), (14, 0), (12, 0), (26, 0), (0, 0)] => Label: priority
[(), 23, 1, (2, 1), (4, 1), (18, 1), (8, 0), (11, 0), (12, 0), (19, 1), (26, 0)] => Label: priority
[(), 25, 0, (26, 0)] => Label: not_recom
[(), 25, 0, (21, 1), (7, 0), (16, 0), (0, 0), (18, 1), (3, 0), (15, 1)] => Label: priority
[(), 22, 1, (15, 1), (25, 1), (9, 0), (0, 0), (6, 1)] => Label: priority
[(), 20, 1, (0, 0), (23, 0), (25, 1), (2, 1), (7, 0), (21, 1), (10, 0), (11, 0), (3, 0)] => Label: priority
[(), 6, 0, (20, 0), (25, 0), (16, 1), (2, 0), (22, 0), (10, 1), (23, 0), (5, 0)] => Label: priority
[(), 13, 0, (15, 0), (11, 1), (16, 1), (23, 0), (5, 1), (26, 0), (1, 1)] => Label: priority
[(), 10, 0, (12, 0), (1, 1), (21, 1), (20, 0), (17, 0), (14, 1), (18, 0), (8, 1), (5, 1), (25, 1)] => Label: priority
[(), 17, 1, (26, 1), (10, 1), (21, 1), (19, 1), (5, 1)] => Label: priority
[(), 2, 1, (6, 1), (10, 1), (21, 1), (25, 1)] => Label: priority
[(), 6, 0, (11, 0), (16, 0), (20, 1), (14, 0), (25, 0), (9, 0), (13, 1), (8, 1), (23, 1), (7, 1), (0, 1), (17, 0)] => Label: priority
[(), 10, 0, (19, 0), (16, 0), (17, 0), (13, 1), (23, 1), (0, 0), (11, 1), (7, 0), (5, 1)] => Label: priority
[(), 1, 1, (26, 1), (22, 1), (16, 1), (3, 0), (20, 1), (5, 1)] => Label: priority
[(), 18, 0, (12, 1), (21, 0), (1, 1), (16, 0), (23, 1), (8, 1), (20, 0), (3, 0), (7, 1), (26, 1)] => Label: priority
[(), 6, 0, (22, 1), (25, 1), (5, 1), (2, 1)] => Label: priority
[(), 13, 0, (16, 0), (20, 0), (26, 0), (14, 0), (1, 1), (12, 0), (8, 0), (18, 1), (10, 1), (21, 1), (5, 1)] => Label: priority
[(), 10, 0, (11, 0), (16, 0), (23, 1), (19, 1), (9, 0), (5, 1), (13, 1), (25, 0), (2, 0)] => Label: priority
[(), 1, 1, (16, 0), (13, 1), (4, 0), (25, 1), (6, 1)] => Label: priority
[(), 19, 0, (14, 1), (16, 1), (2, 1), (7, 0), (10, 0), (8, 0), (22, 0), (11, 1), (26, 1)] => Label: priority
[(), 3, 1, (23, 1), (11, 0), (14, 0), (15, 0), (13, 1), (10, 1), (2, 1), (17, 0), (16, 1), (19, 1)] => Label: priority
[(), 15, 1, (25, 0), (10, 1), (6, 1), (1, 0), (22, 0), (16, 1)] => Label: priority
[(), 11, 1, (5, 1), (16, 1), (15, 0), (20, 0), (23, 0), (2, 0), (0, 1)] => Label: priority
[(), 21, 0, (0, 0), (14, 0), (16, 1), (19, 1), (3, 1), (11, 0), (1, 0)] => Label: priority
[(), 16, 1, (0, 0), (3, 0), (13, 1), (7, 1), (25, 0), (22, 0), (23, 0), (2, 0), (10, 0), (11, 0)] => Label: priority
[(), 5, 1, (12, 1), (19, 1), (1, 0), (1, 0), (22, 0), (0, 1), (18, 0), (23, 0), (17, 0)] => Label: priority
[(), 26, 1, (20, 0), (6, 1), (10, 0), (1, 0), (22, 0), (11, 1), (21, 1), (1, 0), (12, 0)] => Label: priority
[(), 21, 0, (9, 1), (4, 1), (2, 1), (23, 0)] => Label: priority
[(), 10, 0, (14, 0), (11, 0), (18, 1), (0, 1), (26, 0), (13, 1), (6, 1)] => Label: priority
[(), 18, 0, (3, 0), (4, 1), (23, 0), (1, 1), (15, 0), (26, 1), (13, 0), (8, 1), (22, 0)] => Label: priority
[(), 13, 0, (20, 0), (1, 0), (14, 0), (2, 1), (22, 0), (18, 1), (10, 0), (11, 1), (12, 1), (23, 0), (7, 0), (5, 0)] => Label: priority
[(), 20, 1, (5, 1), (25, 0), (23, 0), (10, 1), (1, 1)] => Label: priority
[(), 16, 1, (7, 0), (22, 0), (6, 0), (26, 1), (2, 1), (9, 0), (20, 0), (8, 1), (14, 1), (5, 1)] => Label: very_recom
[(), 10, 0, (7, 1), (17, 1), (25, 0), (19, 0), (23, 1), (9, 0), (8, 1), (1, 0), (15, 0), (12, 1)] => Label: priority
[(), 2, 1, (17, 1), (11, 0), (23, 1), (20, 0), (25, 0), (7, 0), (8, 0)] => Label: priority
[(), 1, 1, (9, 1), (14, 0), (20, 0), (13, 1), (25, 1), (3, 0), (23, 1), (4, 1), (16, 1)] => Label: priority
[(), 13, 0, (0, 0), (20, 1), (11, 1), (3, 0), (5, 1), (17, 1)] => Label: priority
[(), 1, 1, (22, 1), (17, 1), (25, 0), (19, 1), (14, 1), (4, 0), (7, 0), (5, 1)] => Label: priority
[(), 10, 0, (21, 1), (3, 1), (20, 0), (16, 1), (2, 1)] => Label: priority
[(), 25, 0, (16, 0), (18, 0), (13, 1), (2, 1), (20, 0), (11, 1), (21, 0), (5, 1)] => Label: priority
[(), 8, 1, (12, 1), (4, 1), (26, 0), (20, 0), (22, 0), (16, 1)] => Label: priority
[(), 20, 1, (25, 1), (10, 0), (2, 0), (1, 0), (11, 1), (13, 1), (17, 0)] => Label: priority
[(), 19, 0, (7, 0), (4, 0), (0, 0), (22, 0), (13, 1), (3, 0), (25, 0), (6, 1), (10, 0), (2, 1), (8, 0), (17, 0), (9, 1)] => Label: very_recom
[(), 23, 1, (19, 1), (13, 0), (0, 0), (11, 1), (7, 0), (17, 1), (1, 0), (3, 0)] => Label: priority
[(), 14, 1, (6, 0), (5, 1), (23, 0), (1, 1), (25, 1)] => Label: priority
[(), 5, 1, (0, 0), (13, 1), (8, 1), (23, 0), (1, 0), (20, 0), (22, 1)] => Label: priority

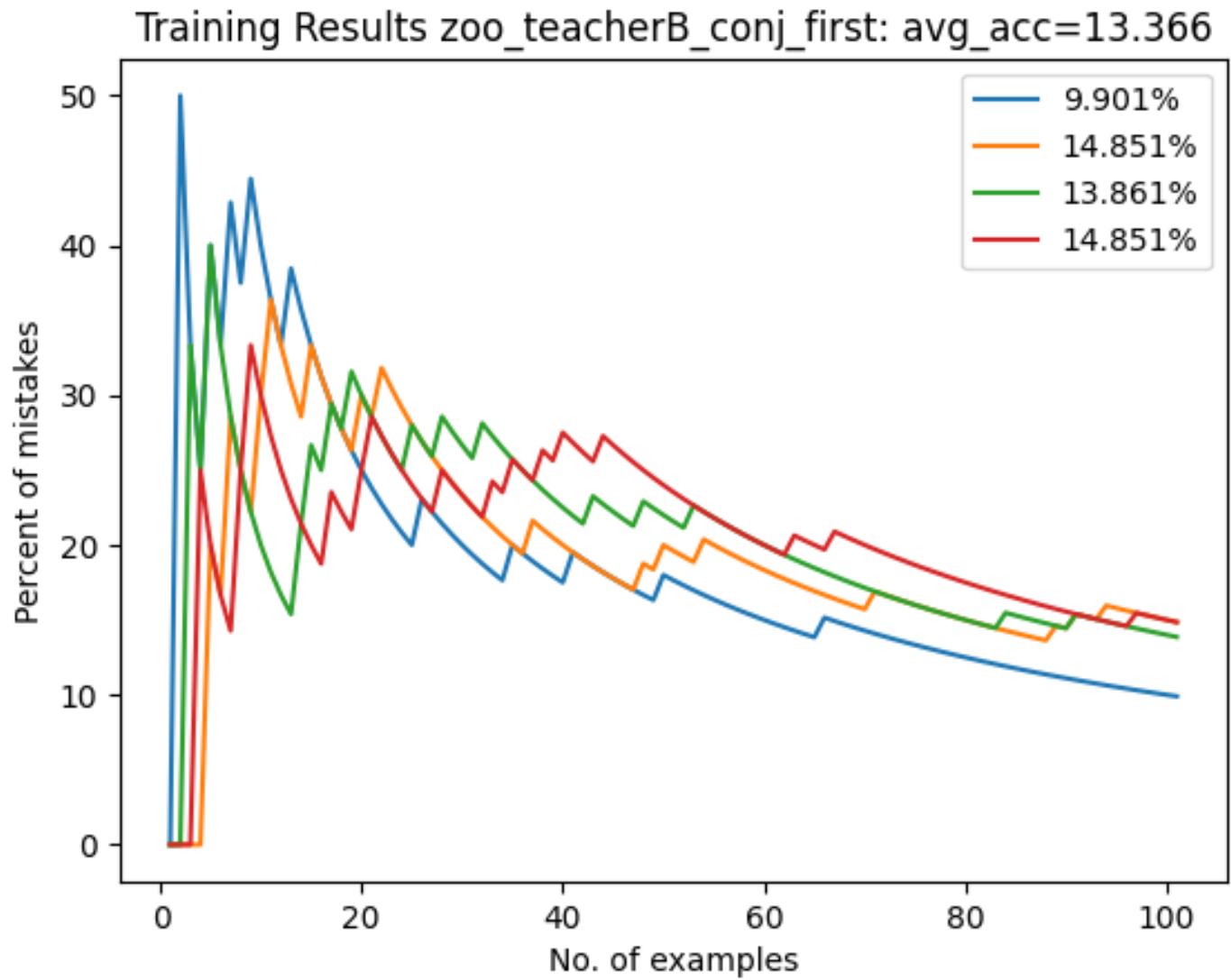
$\{[(18, 0), (11, 0), (17, 1), (6, 1), (15, 1), (9, 1), (23, 1), (25, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(12, 1), (25, 1), (18, 1), (21, 1), (7, 0), (4, 0), (3, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (25, 0), (19, 0), (13, 1), (8, 1), (16, 1), (5, 1), (22, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (26, 1), (9, 0), (0, 0), (19, 0), (2, 1), (8, 1), (7, 0), (12, 1), (3, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(18, 0), (26, 0), (10, 0), (5, 1), (14, 0), (22, 1), (9, 1), (17, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(23, 1), (11, 0), (8, 1), (19, 1), (7, 0), (12, 0), (25, 0), (18, 1), (1, 0), (4, 0), (13, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(13, 0), (9, 1), (20, 0), (6, 1), (21, 0), (1, 1), (12, 1)]\} \Rightarrow \text{Label: very_recom}$
 $\{[(13, 0), (22, 0), (1, 1), (9, 0), (15, 1), (20, 0), (6, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(13, 0), (21, 1), (15, 0), (4, 0), (16, 0), (20, 0), (8, 1), (25, 0), (18, 1), (7, 0), (5, 1), (12, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(10, 0), (12, 0), (17, 0), (25, 0), (9, 1), (23, 1), (15, 0), (18, 1), (20, 0), (14, 0), (6, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(8, 1), (6, 1), (1, 0), (22, 0), (2, 1), (25, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(10, 0), (16, 1), (13, 1), (6, 1), (25, 1), (19, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(8, 1), (4, 1), (14, 0), (20, 1), (17, 1), (12, 1), (1, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (19, 0), (3, 0), (11, 0), (25, 1), (13, 1), (23, 1), (1, 0), (7, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(4, 1), (2, 1), (16, 0), (22, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (12, 1), (25, 0), (9, 1), (16, 1), (1, 0), (2, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(26, 1), (9, 0), (11, 0), (19, 1), (8, 0), (16, 1), (4, 0), (12, 1), (6, 0), (0, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(5, 1), (17, 1), (1, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (18, 1), (9, 1), (19, 0), (3, 0), (4, 0), (7, 0), (5, 0), (13, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(23, 1), (12, 0), (1, 0), (19, 0), (7, 0), (8, 1), (17, 0), (15, 0), (16, 1), (0, 0), (26, 0), (14, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (13, 0), (5, 1), (26, 0), (12, 0), (14, 1), (19, 1), (16, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (8, 1), (0, 1), (22, 0), (7, 0), (15, 1), (3, 0), (5, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(19, 0), (3, 0), (25, 1), (2, 1), (5, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(23, 1), (7, 1), (26, 1), (15, 1), (1, 1), (10, 0), (9, 1), (19, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(3, 1), (11, 1), (21, 0), (23, 0), (16, 1), (25, 1), (20, 0), (2, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(10, 0), (5, 1), (21, 1), (11, 0), (25, 0), (13, 1), (8, 0), (16, 0), (18, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(14, 1), (21, 1), (18, 0), (20, 0), (10, 0), (6, 0), (7, 0), (9, 0), (8, 0), (1, 1), (4, 1), (17, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(18, 0), (23, 0), (20, 0), (11, 1), (26, 0), (16, 1), (7, 0), (13, 1), (0, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(1, 1), (10, 0), (16, 1), (26, 1), (23, 1), (3, 0), (6, 0), (9, 0), (15, 0), (12, 0), (13, 0), (5, 0), (7, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(2, 1), (17, 1), (7, 0), (12, 1), (21, 1), (3, 0), (8, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(2, 1), (15, 0), (4, 0), (21, 1), (20, 0), (10, 1), (16, 1), (3, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (9, 1), (6, 0), (3, 1), (19, 1), (22, 0), (18, 1), (12, 1), (25, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(8, 1), (4, 1), (18, 0), (25, 1), (13, 0), (19, 1), (17, 0), (14, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (22, 0), (18, 1), (6, 1), (14, 1), (0, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(18, 0), (11, 0), (5, 1), (22, 0), (23, 1), (17, 0), (19, 1), (25, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(22, 1), (0, 1), (10, 1), (16, 1), (6, 1), (13, 0), (12, 0), (26, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (14, 0), (21, 0), (15, 0), (3, 0), (10, 1), (25, 0), (22, 1), (2, 1), (4, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(13, 0), (19, 1), (3, 0), (14, 1), (4, 0), (10, 1), (22, 2), (7, 0), (0, 0), (26, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(8, 1), (6, 1), (14, 0), (13, 1), (19, 1), (18, 1), (0, 0)]\} \Rightarrow \text{Label: very_recom}$
 $\{[(10, 0), (22, 0), (17, 0), (7, 1), (25, 0), (18, 1), (2, 0), (21, 1), (1, 0), (12, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(9, 1), (12, 0), (6, 1), (22, 0), (14, 0), (20, 0), (26, 0), (18, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(1, 1), (17, 0), (22, 1), (14, 0), (3, 0), (26, 0), (12, 0), (13, 1), (4, 1), (19, 1), (18, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (3, 0), (15, 0), (13, 1), (17, 0), (22, 1), (5, 1), (0, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(18, 0), (10, 0), (7, 0), (21, 1), (11, 1), (25, 0), (17, 1), (13, 1), (19, 1), (3, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(2, 1), (15, 1), (5, 1), (18, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (20, 0), (0, 0), (1, 0), (5, 1), (23, 0), (18, 1), (22, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (16, 0), (23, 1), (19, 0), (3, 0), (7, 0), (4, 0), (17, 1), (8, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (13, 0), (2, 0), (22, 0), (3, 0), (16, 0), (6, 1), (1, 1), (11, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (9, 1), (13, 1), (18, 1), (21, 1), (0, 1), (4, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(10, 0), (15, 0), (0, 0), (22, 0), (26, 0), (3, 0), (14, 0), (1, 0), (13, 1), (16, 0), (5, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (6, 1), (22, 0), (11, 1), (26, 1), (1, 1), (8, 0), (10, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (20, 0), (12, 1), (2, 0), (4, 0), (1, 1), (18, 1), (3, 1), (21, 1), (8, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(1, 1), (4, 1), (19, 1), (26, 1), (18, 0), (16, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (17, 0), (21, 1), (19, 0), (1, 0), (13, 1), (16, 0), (9, 1), (3, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(2, 1), (18, 1), (7, 0), (20, 1), (11, 0), (14, 0), (3, 0), (4, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(16, 1), (3, 0), (19, 1), (2, 0), (1, 0), (5, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(9, 1), (20, 0), (14, 1), (26, 1), (22, 0), (17, 1), (7, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (1, 0), (12, 1), (8, 1), (22, 0), (6, 0), (2, 0)]\} \Rightarrow \text{Label: priority}$

11, 0), (0, 0), (17, 0), (23, 0), (10, 0), (26, 0), (15, 0), (20, 0), (6, 1)] => Label: priority
5, 0), (7, 0), (3, 1), (1, 1), (16, 1), (14, 0), (10, 0), (15, 0), (12, 1), (22, 0), (26, 1)] => Label: priority
9, 0), (4, 1), (8, 1), (25, 0), (23, 1), (13, 0), (0, 0)] => Label: priority
2, 1), (2, 0), (18, 1), (5, 1), (25, 1), (23, 0)] => Label: priority
0, 1), (22, 0), (25, 0), (15, 0), (8, 1), (17, 0), (13, 1), (1, 0), (0, 1)] => Label: priority
5, 0), (11, 1), (17, 0), (25, 0), (3, 0), (15, 0), (23, 1), (0, 0), (2, 1), (13, 1), (7, 0)] => Label: priority
9, 0), (23, 1), (5, 1), (10, 1), (1, 1)] => Label: priority
5, 1), (2, 1), (17, 1), (4, 1)] => Label: priority
1, 1), (7, 0), (14, 0), (12, 1), (8, 1), (6, 0), (25, 1)] => Label: priority
6, 1), (18, 1), (10, 0), (1, 0), (13, 1), (11, 0), (2, 0), (6, 1)] => Label: priority
8, 0), (1, 1), (20, 1), (25, 0), (16, 1), (14, 0), (9, 0), (4, 0), (13, 1), (3, 0), (22, 1)] => Label: priority
1, (2, 1), (6, 1), (20, 1), (12, 0)] => Label: priority
1, (3, 0), (8, 1), (12, 1), (21, 1), (20, 1), (7, 0)] => Label: priority
6, 1), (21, 1), (0, 1), (3, 0), (11, 0), (19, 1), (8, 1), (17, 1), (4, 0)] => Label: priority
3, 0), (20, 0), (0, 1), (17, 0), (10, 1), (5, 1), (21, 0)] => Label: priority
8, 0), (7, 0), (0, 0), (15, 0), (1, 1), (3, 0), (19, 1), (10, 0), (9, 0), (16, 1), (26, 1), (6, 1)] => Label: very_low
1, (23, 1), (16, 1), (3, 1), (2, 1)] => Label: priority
3, 0), (22, 1), (16, 1), (20, 1), (0, 1), (9, 0), (25, 1), (12, 1), (3, 0), (5, 1), (10, 0)] => Label: priority
5, 0), (22, 0), (0, 0), (10, 1), (7, 0), (1, 0), (18, 1), (3, 0)] => Label: priority
6, 1), (20, 1), (22, 1), (7, 0), (17, 0), (13, 1), (10, 0), (6, 1), (0, 0)] => Label: priority
8, 0), (5, 1), (16, 0), (12, 1), (20, 0), (25, 1), (22, 1)] => Label: priority
2, 1), (1, 1), (4, 1), (10, 0), (23, 0), (9, 0), (19, 0)] => Label: priority
3, 0), (19, 1), (6, 1), (16, 1), (26, 0)] => Label: priority
8, 0), (16, 1), (22, 0), (9, 1), (20, 1), (13, 1), (1, 0), (0, 0), (3, 1)] => Label: priority
1, (4, 1), (25, 1), (21, 1), (15, 0), (1, 1), (18, 1)] => Label: priority
5, 0), (7, 1), (17, 0), (2, 1), (19, 1), (21, 1), (11, 1), (16, 1)] => Label: priority
1, (9, 1), (18, 1), (0, 0), (25, 1)] => Label: priority
1, 0), (4, 0), (6, 1), (25, 0), (2, 1), (20, 0), (11, 1)] => Label: very_recom
2, 1), (25, 1), (19, 1), (17, 0), (14, 0), (3, 1), (8, 1), (13, 1)] => Label: priority
1, 0), (7, 0), (14, 0), (23, 1), (25, 0), (3, 1)] => Label: priority
3, 0), (1, 1), (3, 0), (20, 1), (22, 0), (11, 1), (7, 0), (26, 0), (23, 1), (14, 0)] => Label: priority
0, (20, 0), (13, 0), (15, 1), (22, 0), (11, 0), (5, 1), (1, 0), (25, 1)] => Label: priority
8, 0), (19, 1), (15, 1), (8, 0), (22, 0), (17, 0), (7, 1), (9, 0), (25, 0)] => Label: priority
0, 0), (2, 1), (9, 0), (17, 0), (18, 1), (25, 0), (11, 0), (4, 1)] => Label: priority
8, 0), (26, 1), (20, 1), (21, 1), (5, 1), (8, 0)] => Label: priority
3, 0), (22, 1), (16, 1), (19, 1), (1, 0), (3, 1), (2, 1)] => Label: priority
3, 1), (0, 1), (12, 1), (17, 0), (6, 1)] => Label: priority
8, 0), (2, 1), (22, 0), (12, 1), (25, 0), (3, 1)] => Label: priority
5, 0), (13, 1), (0, 0), (22, 0), (23, 1), (4, 1)] => Label: priority
0, (12, 1), (4, 0), (20, 0), (22, 1), (18, 1), (25, 1), (3, 0), (7, 0)] => Label: priority
7, 1), (11, 0), (9, 1), (21, 0), (22, 0), (13, 0), (0, 0), (7, 0), (25, 0), (15, 0), (4, 1)] => Label: priority
1, 0), (18, 0), (23, 1), (13, 0), (17, 1), (14, 1), (7, 0), (2, 1), (11, 1)] => Label: priority
3, 0), (0, 0), (18, 1), (2, 0), (5, 1)] => Label: priority
5, 0), (4, 1), (26, 0), (14, 1), (11, 0), (0, 0), (8, 0), (19, 0), (23, 0), (21, 1), (16, 1)] => Label: priority
8, 0), (13, 1), (8, 1), (5, 1), (19, 0)] => Label: priority
0, 1), (13, 1), (1, 0), (16, 1), (5, 1), (9, 1)] => Label: priority
9, 0), (2, 1), (22, 0), (12, 1), (17, 0), (25, 0), (4, 0), (10, 0), (7, 1)] => Label: priority
0, 0), (16, 1), (2, 1), (25, 0), (15, 0), (19, 1), (12, 0), (11, 1)] => Label: very_recom
4, 1), (1, 0), (4, 0), (20, 0), (21, 0), (10, 0), (2, 0), (16, 1)] => Label: priority
0, (22, 0), (23, 1), (2, 1), (8, 0), (11, 0), (12, 1), (26, 1)] => Label: priority
0, 0), (20, 1), (11, 0), (12, 1), (22, 0), (26, 1), (3, 0), (17, 0), (0, 1)] => Label: priority
6, 1), (17, 0), (16, 1), (14, 0), (0, 1), (8, 0), (11, 1), (22, 0)] => Label: priority
1, 1), (0, 1), (26, 1), (18, 1), (5, 1), (13, 1), (19, 1)] => Label: priority
1, 0), (25, 0), (21, 1), (15, 0), (14, 0), (1, 1), (12, 0), (18, 1), (7, 1)] => Label: priority
5, 0), (16, 0), (1, 0), (19, 1), (0, 0), (21, 1), (7, 0)] => Label: priority
0, 12, 1), (23, 1), (16, 1), (1, 1), (9, 0), (10, 1), (7, 0)] => Label: priority
0, 0), (11, 0), (20, 0), (22, 0), (5, 1), (8, 1), (13, 1)] => Label: priority
1, 0), (23, 0), (15, 0), (10, 1), (13, 1), (3, 0), (7, 0), (6, 1), (26, 1), (17, 0)] => Label: priority

[(), 0, (8, 1), (7, 0), (25, 0), (22, 0), (20, 1), (18, 1), (5, 0)] => Label: priority
[(25, 0), (0, 0), (9, 1), (19, 0), (4, 1), (2, 0)] => Label: priority
[(2, 1), (15, 0), (20, 1), (12, 0), (14, 0), (16, 1), (11, 0), (21, 0), (25, 1), (22, 1), (7, 0), (9, 1)] => Label: priority
[(26, 1), (0, 0), (20, 0), (15, 0), (22, 0), (12, 1), (7, 1), (11, 0), (17, 0)] => Label: priority
[(10, 0), (17, 0), (20, 0), (8, 1), (13, 1), (16, 0), (23, 1), (25, 0)] => Label: priority
[(10, 0), (5, 1), (17, 0), (26, 0), (1, 1)] => Label: priority
[(20, 1), (21, 0), (2, 1), (4, 1), (12, 0)] => Label: priority
[(16, 1), (7, 0), (1, 1), (22, 0), (3, 0), (10, 1), (20, 1), (6, 1)] => Label: priority
[(6, 0), (3, 0), (4, 0), (16, 1), (11, 0), (25, 0), (0, 0), (2, 1), (21, 1), (20, 0)] => Label: priority
[(2, 1), (23, 1), (7, 0), (8, 0), (12, 1), (10, 0)] => Label: priority
[(6, 0), (16, 1), (3, 0), (7, 0), (11, 0), (22, 0), (10, 1), (15, 0), (5, 0), (25, 0)] => Label: priority
[(10, 0), (6, 1), (25, 0), (15, 0), (23, 1), (16, 0)] => Label: priority
[(9, 1), (21, 1), (25, 0), (18, 1), (20, 0), (3, 1)] => Label: priority
[(2, 1), (23, 0), (21, 1), (9, 0), (7, 0), (4, 1), (25, 1)] => Label: priority
[(11, 1), (25, 0), (23, 1), (18, 0), (0, 0)] => Label: priority
[(13, 0), (26, 1), (9, 1), (12, 1), (18, 0), (22, 1), (0, 0), (5, 1)] => Label: priority
[(25, 0), (7, 0), (22, 0), (4, 0), (5, 1), (10, 0), (11, 1)] => Label: priority
[(18, 0), (8, 1), (16, 1), (19, 1), (22, 1), (7, 0), (0, 1), (25, 1)] => Label: priority
[(1, 1), (22, 1), (20, 0), (7, 0), (8, 0), (5, 1)] => Label: priority
[(9, 1), (7, 0), (2, 1), (20, 1), (12, 1), (16, 1)] => Label: priority
[(7, 1), (0, 1), (20, 0), (8, 1), (21, 1)] => Label: priority
[(10, 0), (21, 1), (0, 0), (25, 1), (13, 1), (16, 1), (9, 1), (7, 0)] => Label: priority
[(8, 1), (5, 1), (26, 1)] => Label: priority
[(20, 1), (18, 1), (23, 1), (26, 1), (10, 0), (13, 0), (12, 1)] => Label: priority
[(26, 1), (0, 0), (17, 0), (14, 0), (18, 0), (1, 0), (22, 0), (23, 0), (15, 0)] => Label: very_recom
[(25, 0), (18, 1), (10, 1), (21, 1), (1, 0), (20, 0), (7, 0)] => Label: priority
[(13, 0), (25, 1), (20, 0), (14, 1), (8, 1), (7, 0), (1, 1), (22, 0)] => Label: priority
[(6, 0), (23, 1), (5, 1), (0, 0), (26, 1)] => Label: very_recom
[(6, 0), (3, 0), (15, 1), (5, 1), (0, 0)] => Label: very_recom
[(11, 1), (25, 0), (22, 1), (4, 0), (18, 1), (15, 1), (7, 0)] => Label: priority
[(13, 0), (10, 0), (12, 1), (8, 1), (3, 0), (17, 0), (5, 1)] => Label: priority
[(3, 1), (9, 0), (23, 1), (19, 1), (0, 1), (17, 0), (10, 1)] => Label: priority
[(22, 1), (5, 0), (16, 0), (14, 0), (6, 1), (19, 1), (15, 1), (10, 0)] => Label: priority
[(6, 0), (23, 1), (16, 1), (4, 1), (19, 1), (14, 0)] => Label: priority
[(19, 0), (7, 0), (16, 0), (2, 1), (8, 0), (22, 1), (25, 1), (11, 0), (13, 1)] => Label: priority
[(8, 1), (5, 0), (26, 0), (23, 0), (4, 1), (0, 0)] => Label: priority
[(1, 1), (3, 0), (4, 0), (15, 1), (7, 0)] => Label: priority
[(6, 0), (25, 1), (13, 1), (8, 1), (18, 1), (0, 1), (19, 0)] => Label: priority
[(11, 1), (7, 0), (17, 0), (3, 0), (15, 0), (21, 0), (12, 1)] => Label: priority
[(13, 0), (15, 1), (17, 0), (11, 0), (16, 1), (21, 1), (7, 0), (1, 0), (10, 0), (2, 1)] => Label: priority
[(20, 1), (0, 1), (6, 1), (25, 1), (23, 0)] => Label: priority
[(26, 1), (21, 0), (19, 1), (23, 1), (6, 1)] => Label: priority
[(10, 0), (18, 1), (12, 1), (4, 0), (3, 1)] => Label: priority
[(0, 0), (8, 1), (1, 0), (22, 1), (3, 1)] => Label: priority
[(11, 1), (26, 1), (13, 1), (21, 1)] => Label: priority
[(6, 0), (9, 0), (26, 0), (20, 0), (0, 0), (11, 0), (13, 0), (3, 0)] => Label: priority
[(19, 0), (17, 0), (2, 1), (15, 0)] => Label: priority
[(8, 1), (3, 0), (0, 0), (4, 1)] => Label: priority
[(6, 0), (15, 0), (14, 0), (25, 0), (21, 0)] => Label: priority
[(12, 1)] => Label: priority

Part A

- Zoo:
- 4 runs of TeacherB



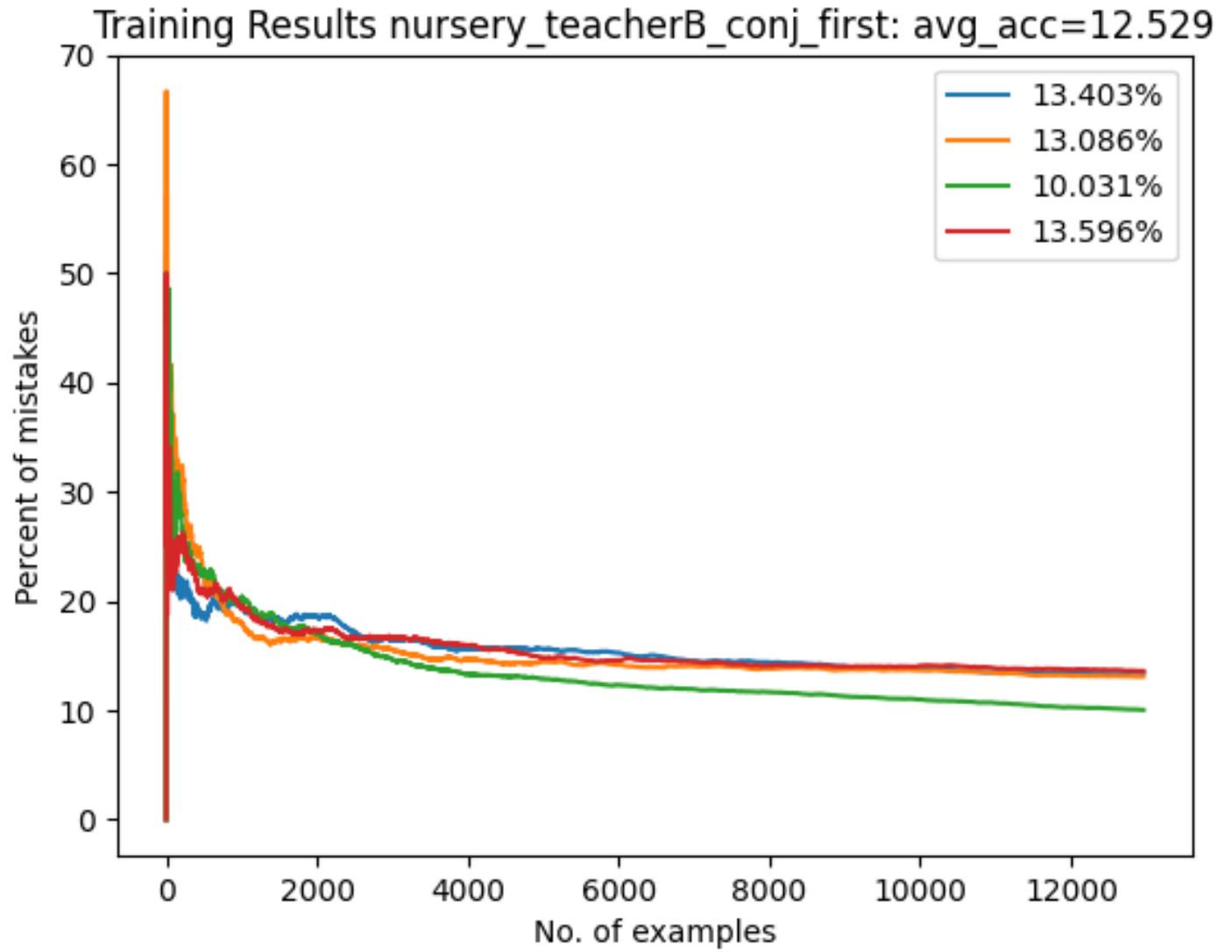
Part A

- Zoo:
- Example of a final decision list TeacherB

```
[ (3, 1) ] => Label: 1
[ (1, 1) ] => Label: 2
[ (9, 1), (5, 1) ] => Label: 5
[ (7, 0), (19, 0), (8, 0) ] => Label: 7
[ (11, 0), (8, 1) ] => Label: 3
[ (5, 0) ] => Label: 6
```

Part A

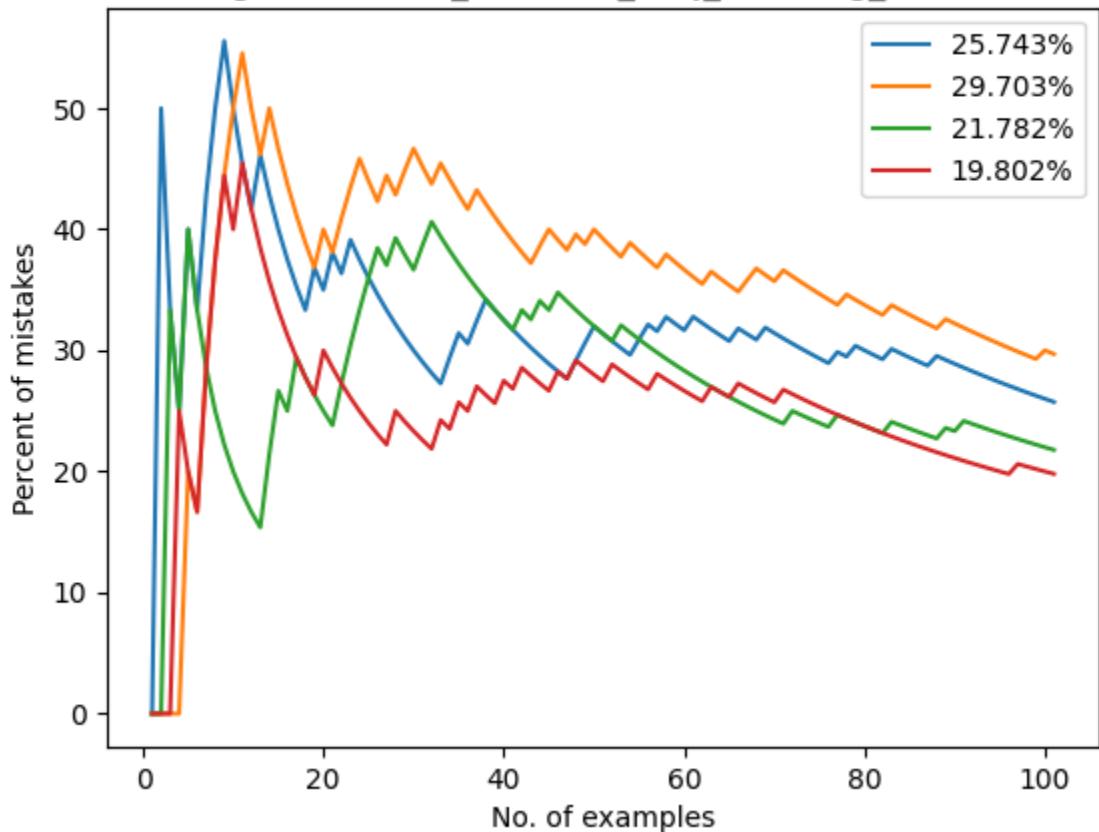
- Nursery:
- 4 runs of TeacherB



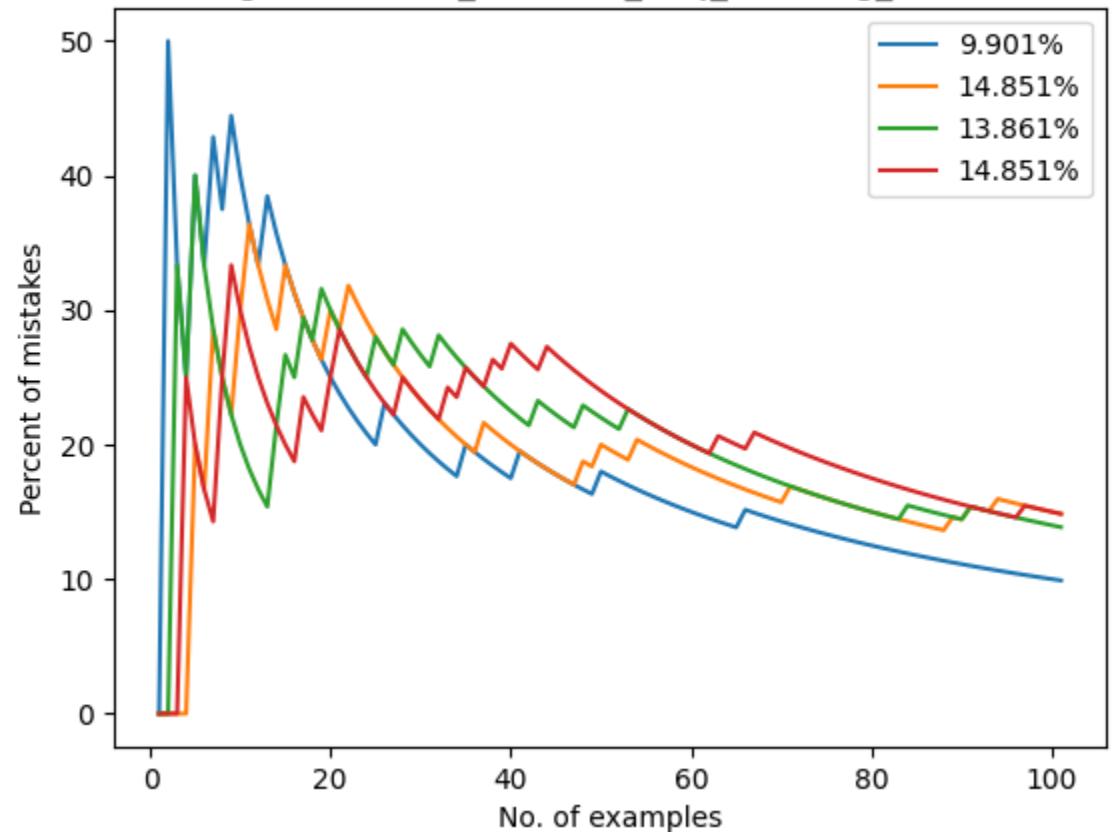
$\{[(0, 0), (3, 0), (7, 0), (25, 0), (22, 0), (16, 1), (12, 0), (2, 0), (6, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (7, 0), (2, 1), (16, 1), (14, 0), (11, 0), (8, 0), (22, 1), (25, 1), (15, 0), (10, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (0, 1), (25, 0), (16, 1), (21, 0), (22, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (25, 0), (2, 1), (19, 0), (16, 0), (21, 1), (15, 0), (8, 0), (17, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (0, 0), (12, 1), (7, 0), (2, 1), (10, 0), (17, 0), (22, 0), (21, 1)]\} \Rightarrow \text{Label: very_recom}$
 $\{[(0, 0), (25, 0), (7, 0), (2, 0), (17, 0), (22, 0), (16, 0), (19, 0), (8, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (7, 0), (2, 1), (3, 0), (21, 1), (16, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(5, 1), (22, 0), (19, 0), (16, 0), (11, 0), (17, 0), (25, 1), (12, 0), (10, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (2, 0), (0, 0), (16, 1), (8, 1), (25, 0), (22, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (6, 1), (17, 1), (22, 1), (19, 0), (12, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(22, 1), (16, 0), (0, 1), (6, 1), (8, 1), (17, 0), (25, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (3, 1), (0, 1), (25, 0), (15, 0), (16, 0), (12, 0), (22, 0), (17, 0), (14, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (2, 1), (7, 0), (22, 0), (25, 0), (12, 1), (16, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (7, 0), (3, 1), (0, 1), (16, 1), (12, 1), (22, 1), (19, 1), (8, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (3, 0), (17, 0), (7, 0), (25, 1), (22, 1), (16, 0), (15, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (25, 0), (0, 0), (2, 1), (22, 0), (16, 1), (21, 1), (8, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (0, 0), (16, 0), (2, 0), (22, 0), (25, 0), (17, 0), (19, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (7, 0), (25, 0), (3, 0), (16, 1), (22, 0), (12, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(22, 1), (0, 1), (6, 1), (17, 0), (25, 1), (19, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (25, 0), (21, 1), (0, 0), (2, 1), (17, 0), (14, 0), (10, 0), (15, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (7, 0), (2, 1), (10, 0), (12, 1), (17, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (3, 0), (7, 0), (17, 0), (25, 1), (16, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(16, 1), (7, 0), (5, 0), (6, 1), (8, 1), (15, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(5, 1), (19, 0), (25, 1), (22, 1), (8, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(6, 0), (7, 0), (3, 1), (0, 1), (17, 1), (22, 0), (12, 1), (25, 0)]\} \Rightarrow \text{Label: very_recom}$
 $\{[(25, 0), (0, 0), (22, 0), (16, 0), (7, 0), (12, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(21, 0), (7, 0), (6, 1), (25, 1), (16, 0), (22, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (25, 0), (0, 0), (17, 0), (16, 0), (2, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (0, 0), (2, 1), (16, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (2, 1), (3, 0), (7, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(0, 0), (2, 1), (17, 0), (7, 0)]\} \Rightarrow \text{Label: priority}$
 $\{[(7, 1), (25, 0), (22, 0), (0, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(25, 0), (0, 0), (7, 0), (22, 0), (21, 1)]\} \Rightarrow \text{Label: very_recom}$
 $\{[(5, 1), (25, 0), (19, 1)]\} \Rightarrow \text{Label: priority}$
 $\{[(5, 1)]\} \Rightarrow \text{Label: priority}$

Conclusions

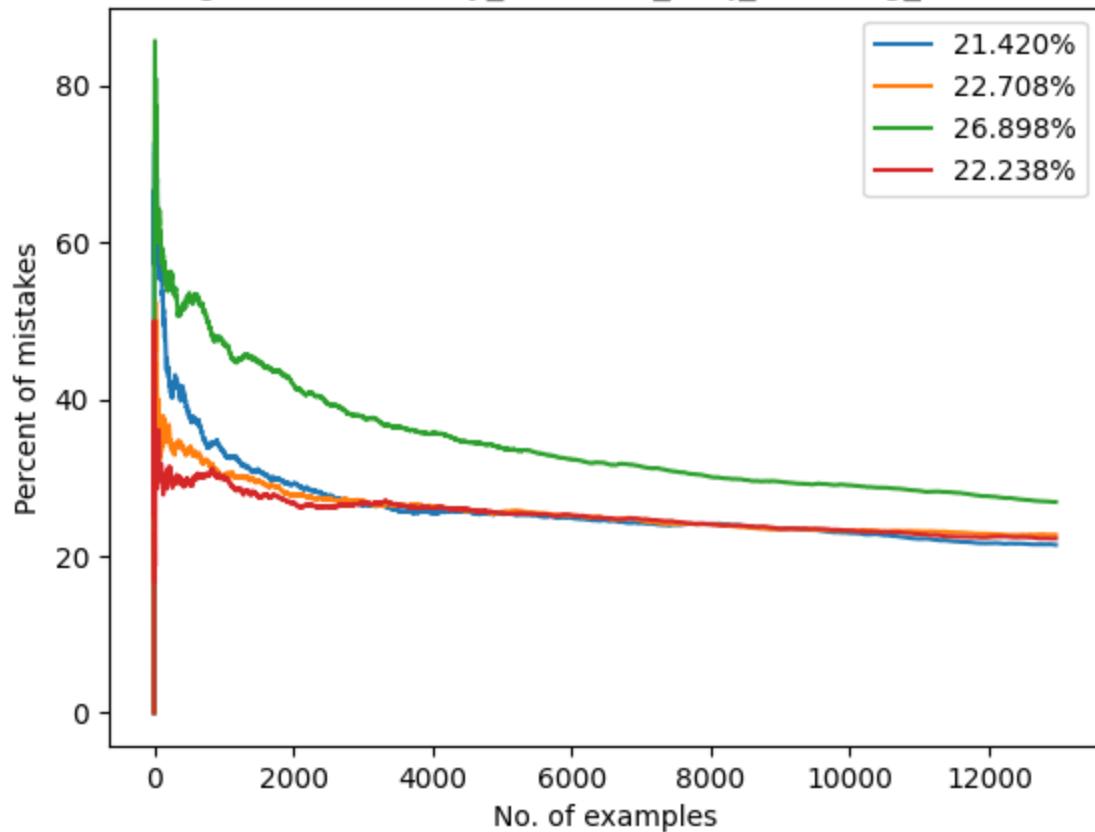
Training Results zoo_teacherA_conj_first: avg_acc=24.257



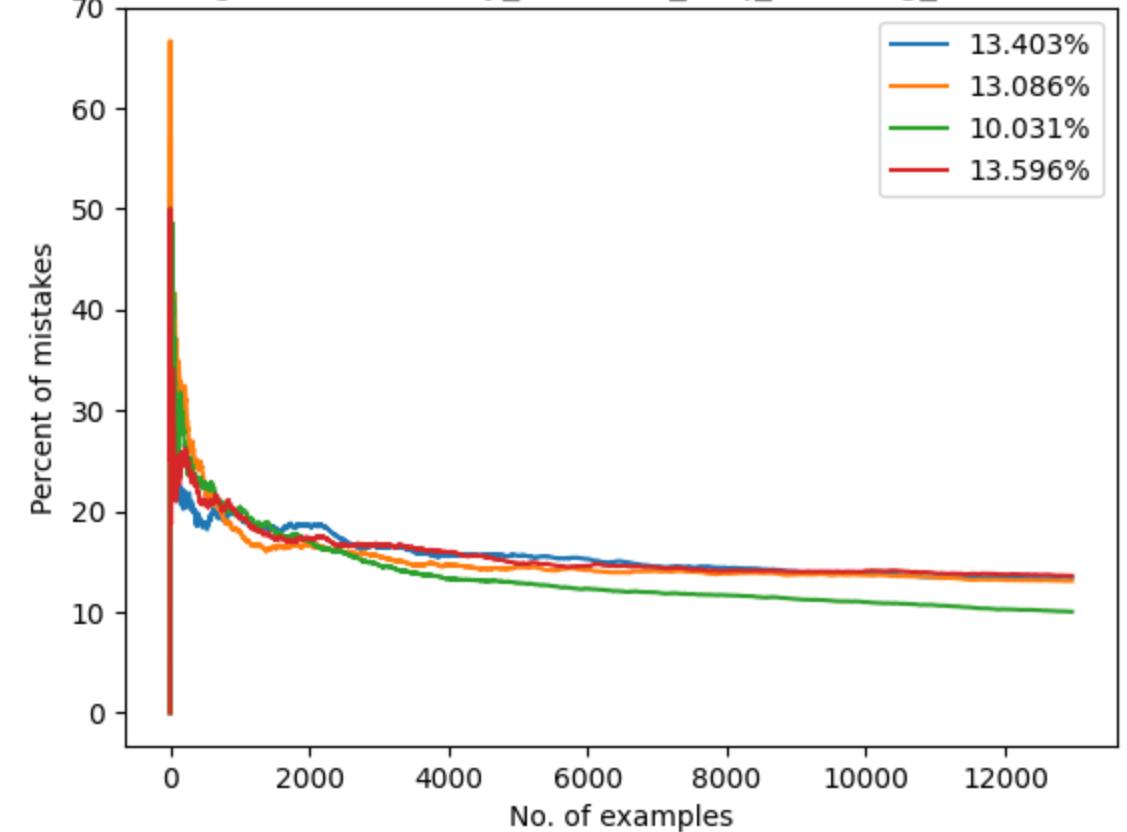
Training Results zoo_teacherB_conj_first: avg_acc=13.366



Training Results nursery_teacherA_conj_first: avg_acc=23.316



Training Results nursery_teacherB_conj_first: avg_acc=12.529



Conclusions

- In both the zoo and the nursery data base we can see teacherB is a lot better in accuracy(9% better)
- since most animals in a certain label have many similar features teacherB outputs better accuracy as it looks on each label as one similar group and its stats are quite representative.
- if each animal group was divided to many unalike sub groups then maybe teacherA would have been better.
- In both datasets the TeacherB runs give smaller mistake percentages and shorter decision lists
- Also, in the TeacherB runs, we can see in the graphs that the decline in the relative mistake percentage begins sooner in the run of the algorithm.
- choosing the discriminative feature in random is a problem for the algorithm as it will not be discriminative in *high probability this will cause* the conjunction group to not identify the item in an optimal way leading to more prediction misses.
- Overall, TeacherB improves the accuracy of the decision list. Such that each conjunction better represents a subgroup of a label.

Part B

Our Dataset

- For this part we chose the [Mushroom Dataset](#)
- It had 8124 instances and 22 attributes (127 after the binary conversion)
- This data set includes descriptions of hypothetical samples corresponding to 23 species mushrooms.
- Each attribute represents a feature of the sample

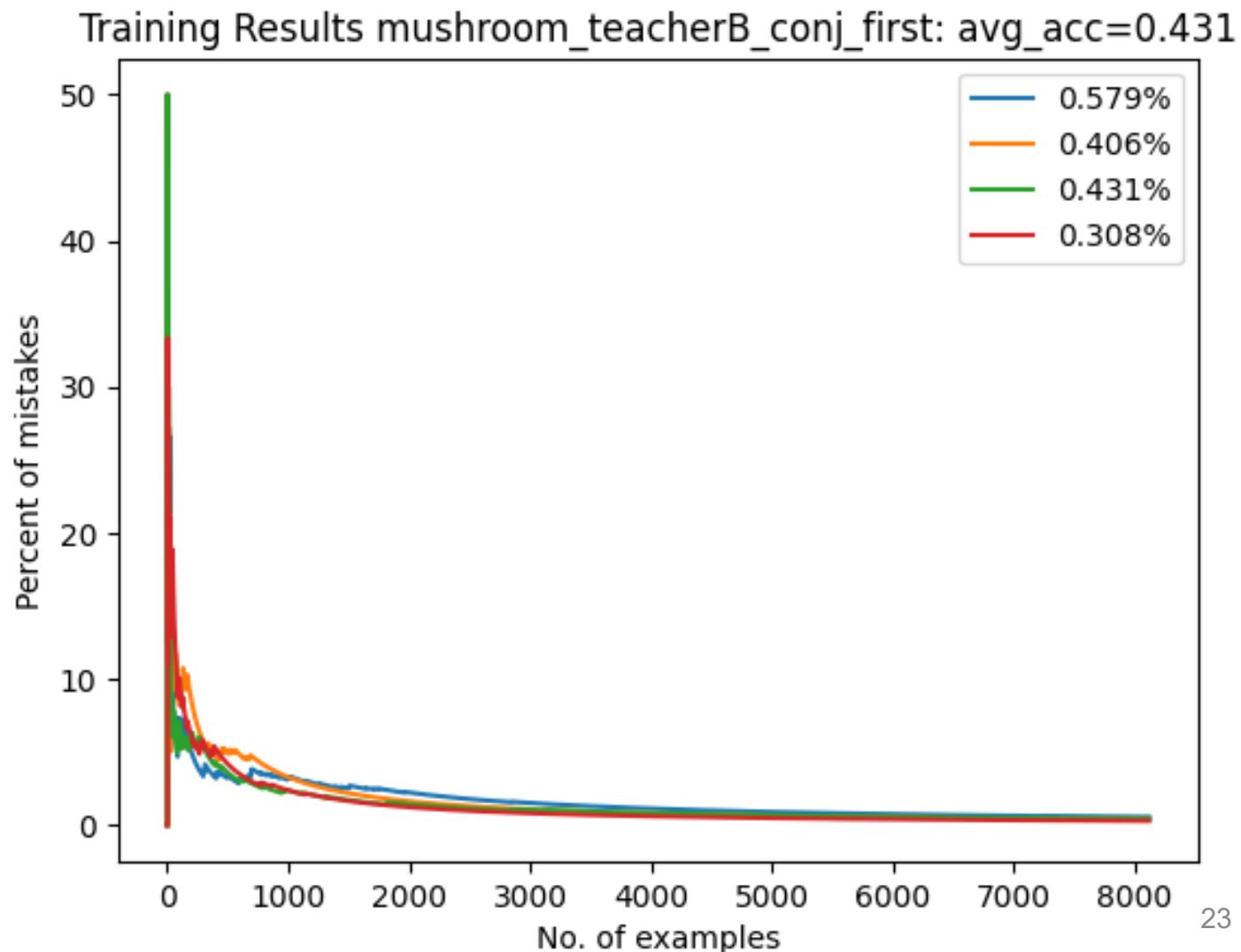


Attribute Information:

- 1. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
- 2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
- 3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y
- 4. bruises?: bruises=t,no=f
- 5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
- 6. gill-attachment: attached=a,descending=d,free=f,notched=n
- 7. gill-spacing: close=c,crowded=w,distant=d
- 8. gill-size: broad=b,narrow=n
- 9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y
- 10. stalk-shape: enlarging=e,tapering=t
- 11. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
- 12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
- 13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
- 14. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
- 15. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
- 16. veil-type: partial=p,universal=u
- 17. veil-color: brown=n,orange=o,white=w,yellow=y
- 18. ring-number: none=n,one=o,two=t
- 19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
- 20. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y
- 21. population: abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y
- 22. habitat: grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d

- The instances are divided to 2 classes (labels): edible or poisonous
- 4208 (51.8%) are edible and 3916 (48.2%) are poisonous
- We chose this dataset for its small set of values to each attribute

TeacherB on Mushrooms



Our ideas to improve the algorithm

- 1) improve the teacher to identify better what is the most discriminative feature for each item.
- 2) improve the learner so it picks a better conjunction for the prediction the label
- 3) quality of life control. Such as: deleting a predicate in the conjunction if the number of predicates is too long

1) The Teachers

- We implemented three additional teachers:
- 1) TeacherC – finds a discriminative feature comparing the current instance with an average of the instances that represent the conjunction
- 2) TeacherD – use the feature statistics from the potential instances that satisfy the conjunction predicates and label
- 3) TeacherE – divides the label to subgroups using a decision tree and compares the conjunctions to those subgroups

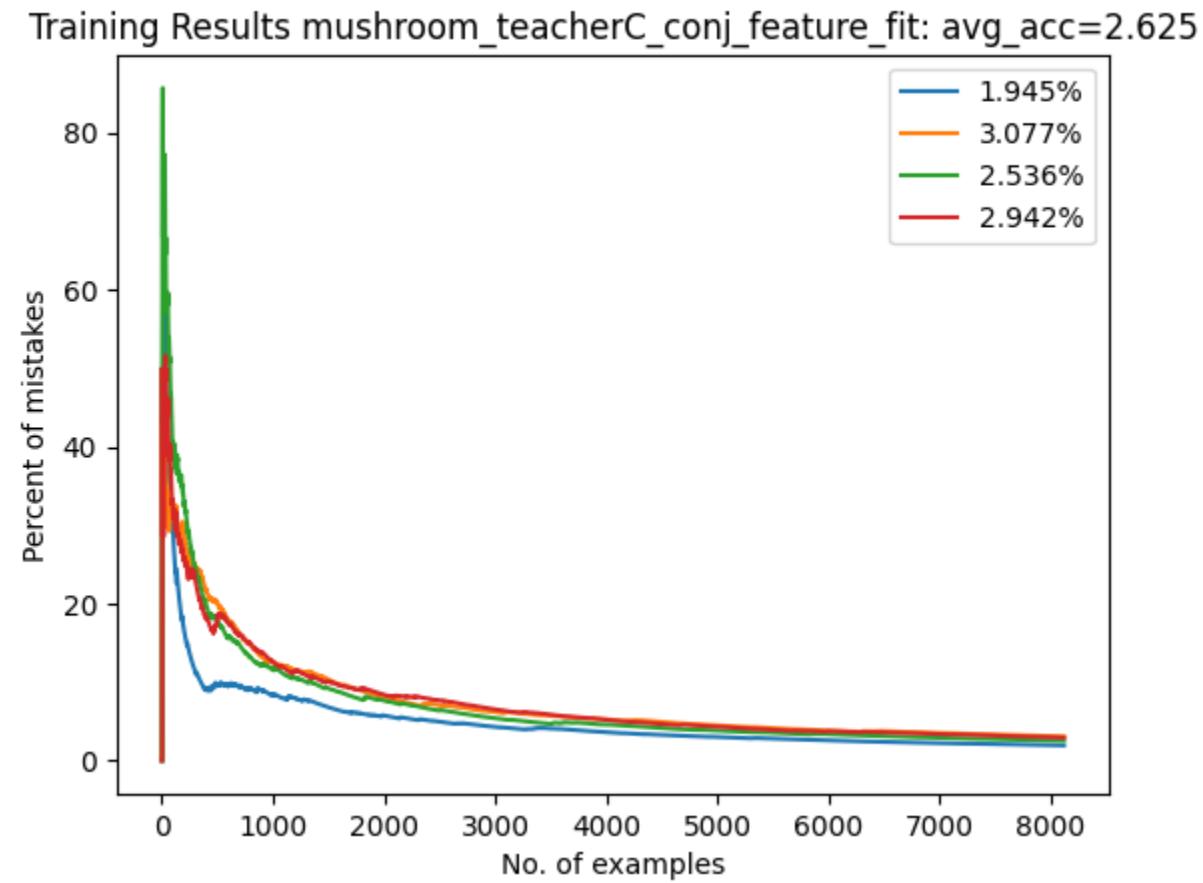
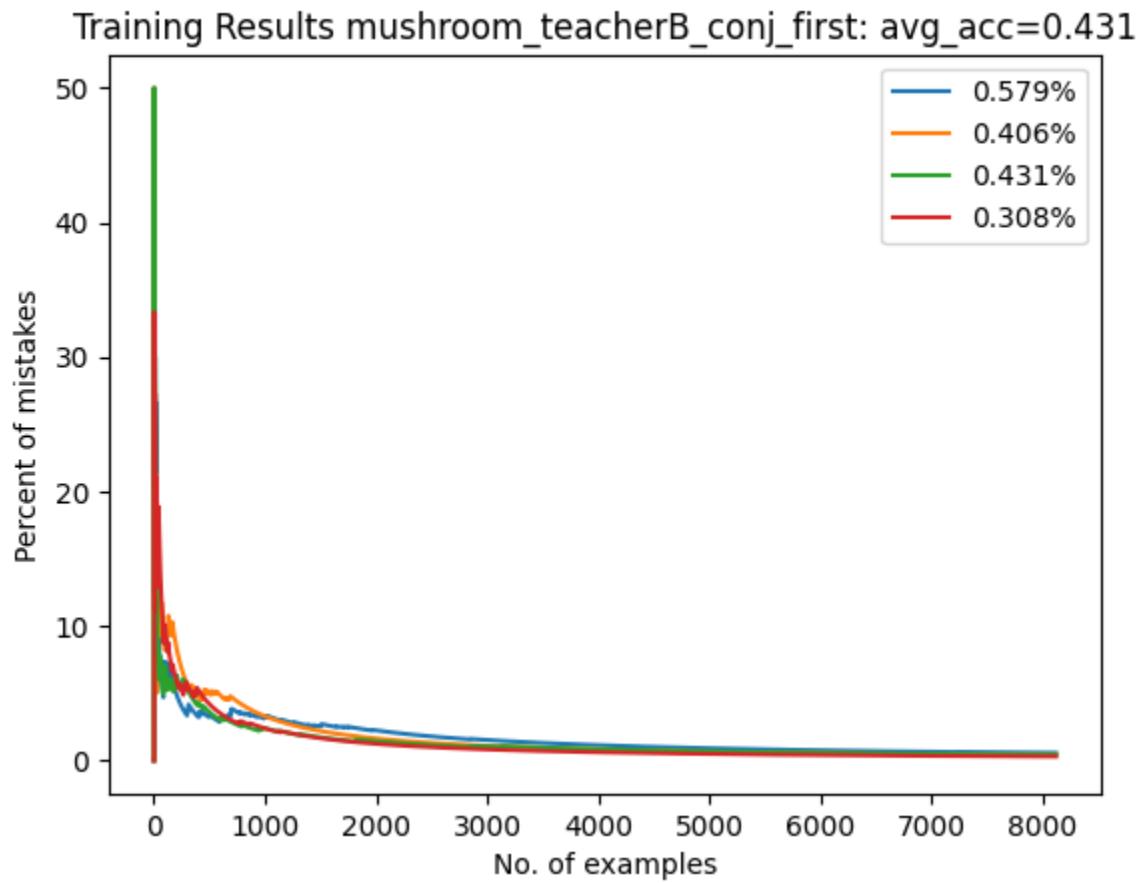
MOTIVATION

- The problem: in TeacherB, each time we compare two instances we compare their entire labels.
- looking at the entire group statistics might result in less optimal discriminative features than looking at only a sub group
- if the dataset has a more uniform feature distribution in each label there might be potential "sub groups" that are more alike
- Solution: create smaller subgroups from the label group and compare those groups percentages instead.

TeacherC

- we will save on each conjunction the instances that predicted correctly on said conjunction as a new subgroup of the label.
- We will compare the groups normalized feature list (%) to the predicted instance.
- When we compare, we compare $\text{abs}(\text{normalized_conjunction}-\text{predicted_instance_features})$ and we pick the feature with the highest score.

Results – TeacherC



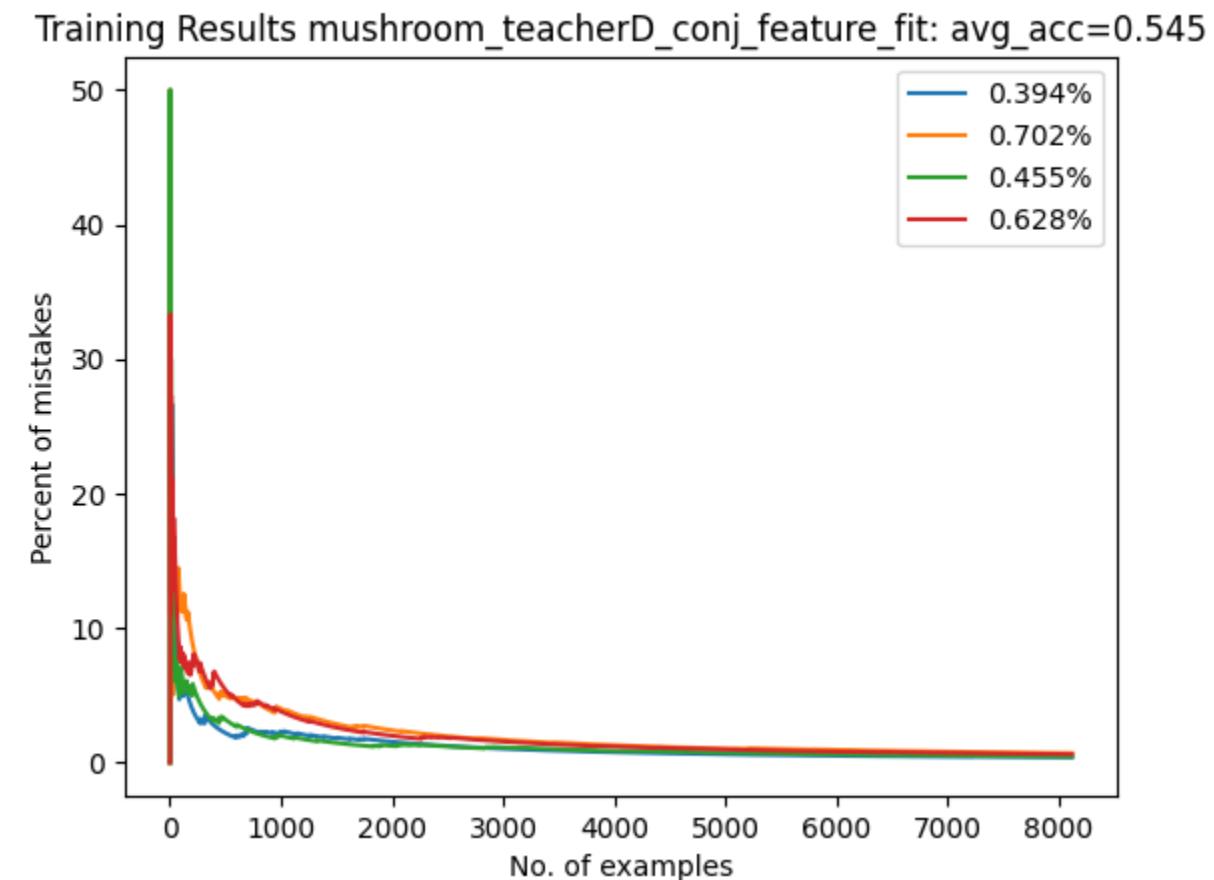
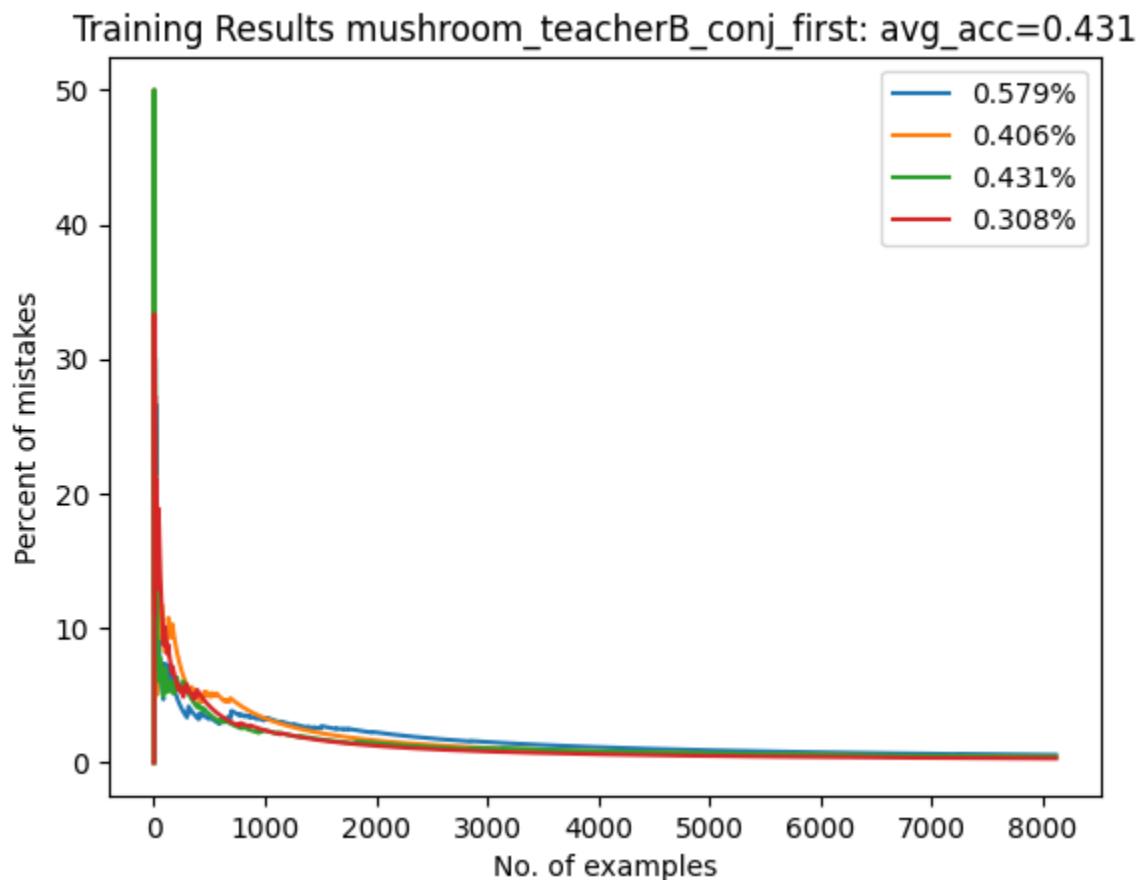
Conclusion

- On the Mushroom dataset TeacherC was incorrect on average 2.625% of the times.
- We can see that the algorithm running with TeacherB on the Mushroom Dataset brought better and more accurate results.
- Trying to solve the problem we presented with TeacherC did not improve the accuracy.

TeacherD

- Here instead of looking at all the label group statistics we choose only items that satisfy the current conjunction predicates.
- This is another attempt at a comparison different than the one in TeacherB
- Here we compare:
 $\text{abs}(\text{normalized_conjunction} - \text{normalized_predicted_label_group})$
and we pick the feature with the highest score.

Results - TeacherD



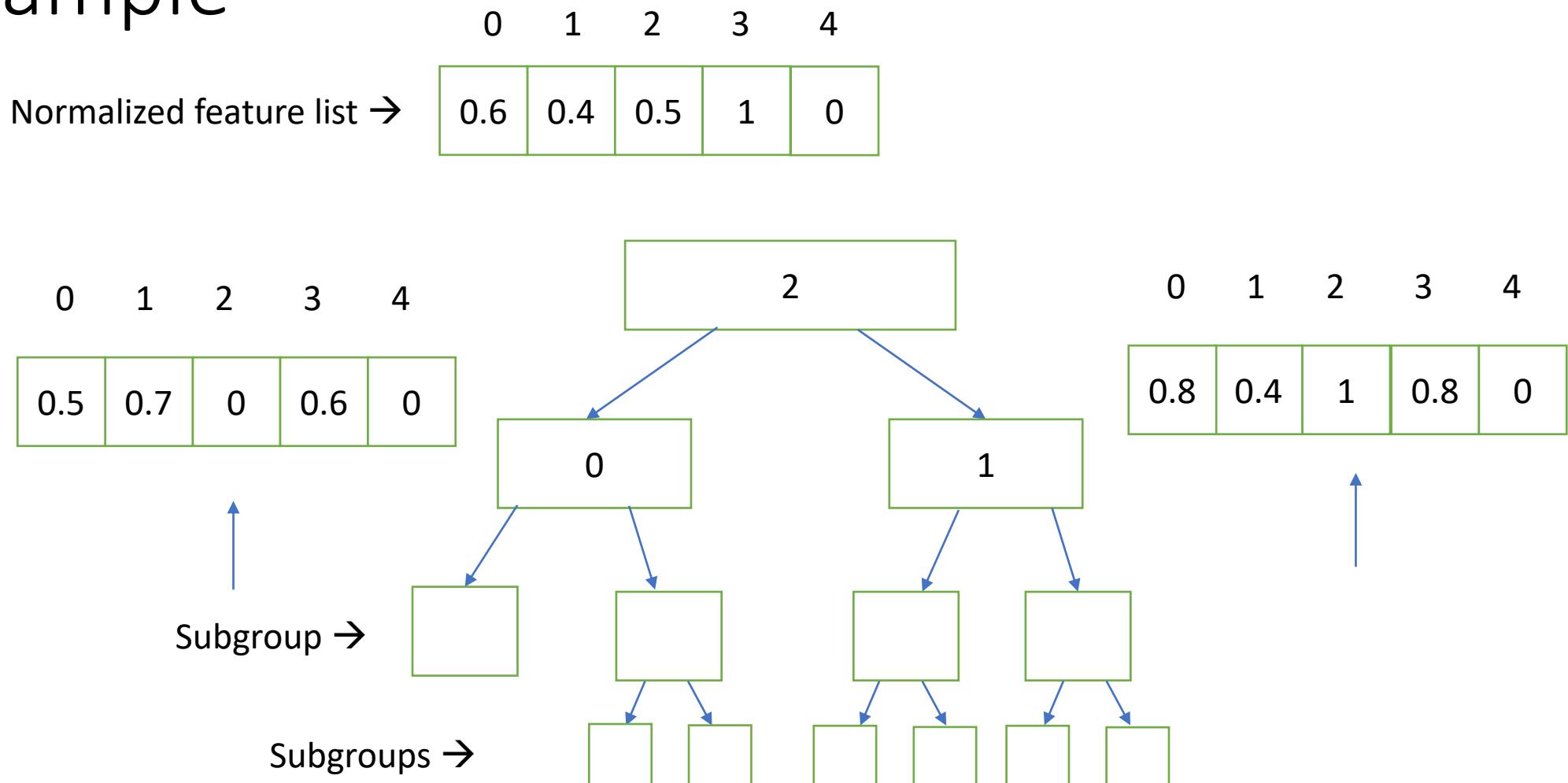
Conclusion

- On the Mushroom dataset TeacherD was incorrect on average 0.545% of the times.
- TeacherD brought similar results to TeacherB on the Mushroom Dataset
- Trying to compare the normalized conjunction to the entire label instead of the single instance brought better results.
- This Teacher is more accurate than TeacherC on the Mushroom Dataset

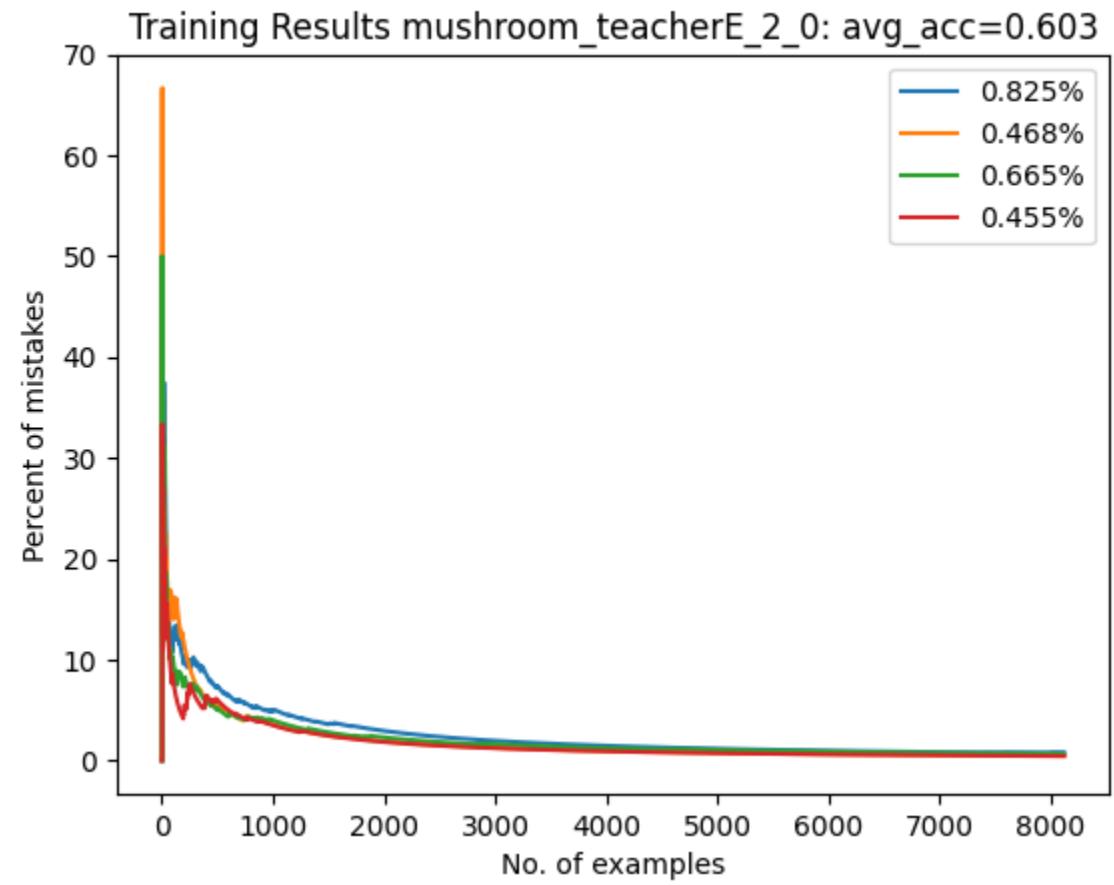
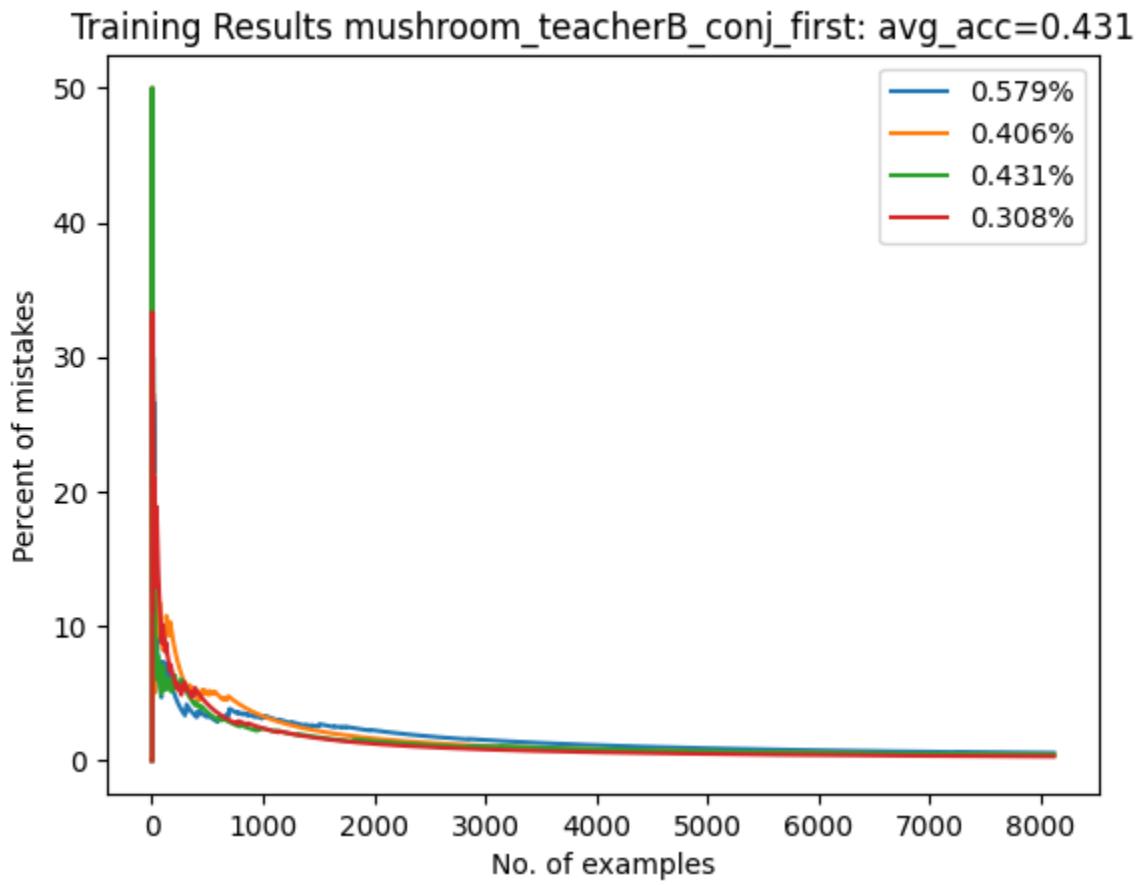
TeacherE

- We decided to create subgroups for each label using a classifier.
- We create a decision tree for each label based on the normalization of all the instances in the group.
- In the initialization of the teacher, we calculate the normalized feature list of all the instances in each label.
- From this list we take the features with values that are in a pre-set range of values (always around 0.5). We then go in order of the features with the values that are closest to 0.5 and for each one create a level of nodes in the tree.
- each branch node in the tree represents a feature. the right son of the node has all items with the feature enabled and the left son has all items with the features disabled. the leaves of the tree represent the subgroups of the label.
- We also make sure the tree doesn't create too much levels and too small subgroups.
- The tree is created top-down to match criteria it was given (percentage_from_mid, max_tree_level, min_group_size).
- During the run of the learning algorithm, when the teacher wants to get statistics on an item, it goes from the top of the tree down, and selects left or right according to the node in the feature and its value in the item. when it gets to a leaf node, that node has the subgroup of the item and its statistics
- The teacher compares:
$$\text{abs}(\text{normalaized_tree_sub_group_of_conj_instance} - \text{normalaized_tree_sub_group_of_instance})$$

Example

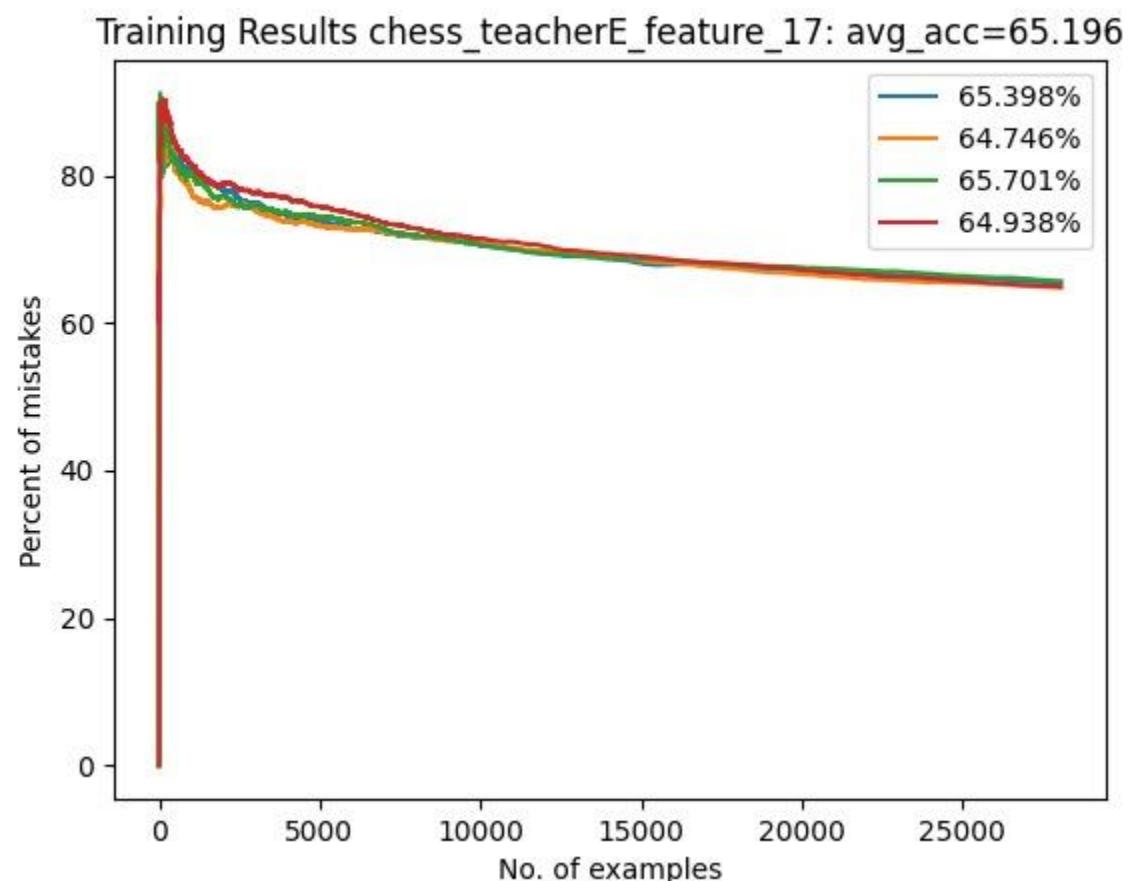
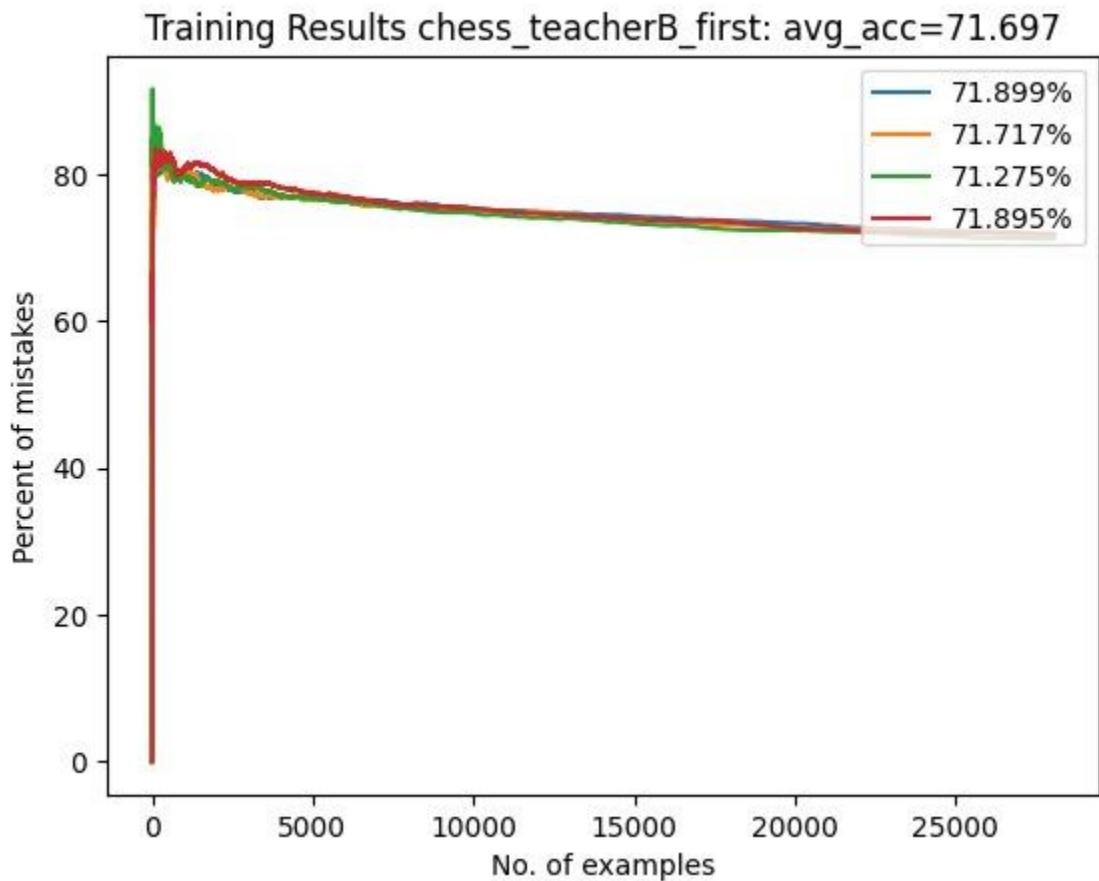


Results



Other Datasets

On the Chess dataset there was an improvement of 6.5%



Conclusion

- On the Mushroom dataset TeacherE was incorrect on average 0.603% of the times.
- TeacherE on the Mushroom dataset had similar results to TeacherB.
- TeacherE did not improve on the Mushroom dataset but did improve on the Chess dataset in a significant way (6.5%).
- Overall, teacherE is random so it might not pick the best groups. maybe the correlation between all the items in the larger dataset is smaller so teacherE is more effective even if it is random

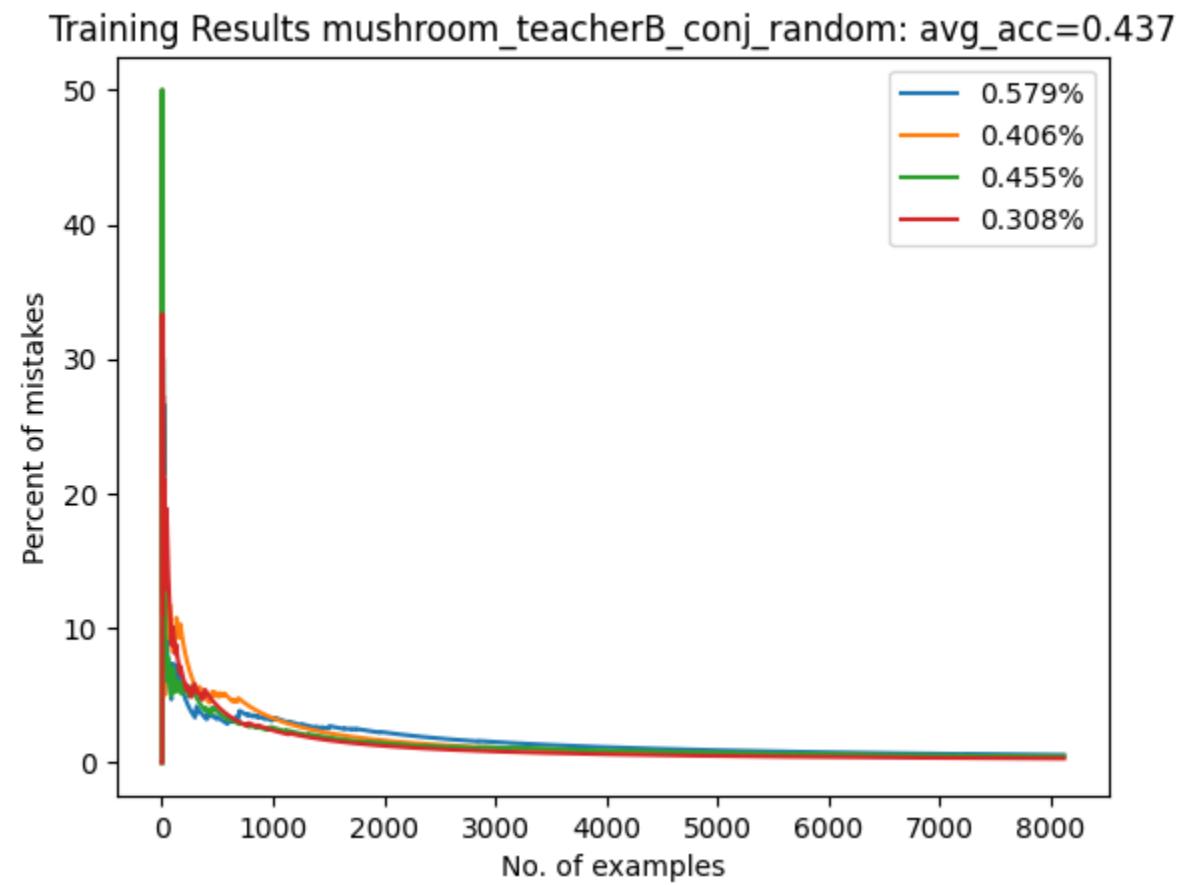
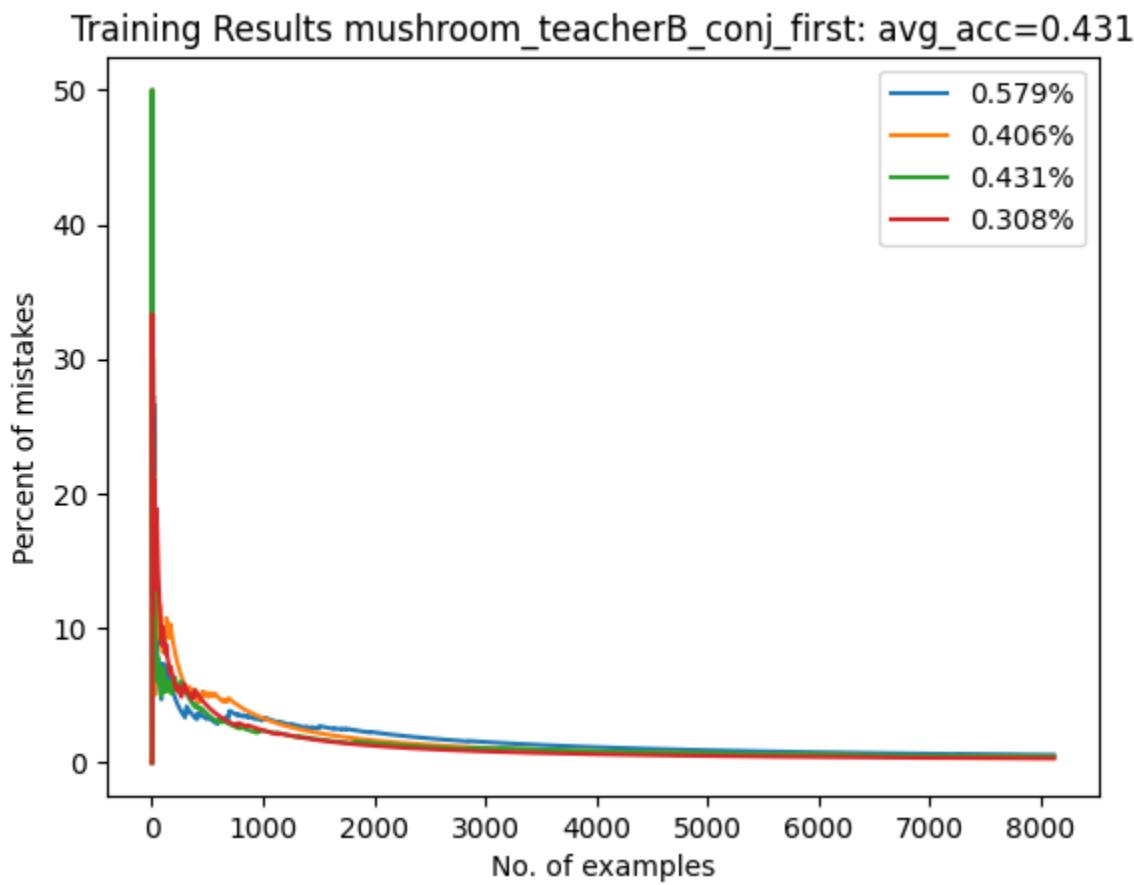
2) MOTIVATION

- Picking the right conjunction can influence how the algorithm learns and his accuracy.
- We implemented a class for each method

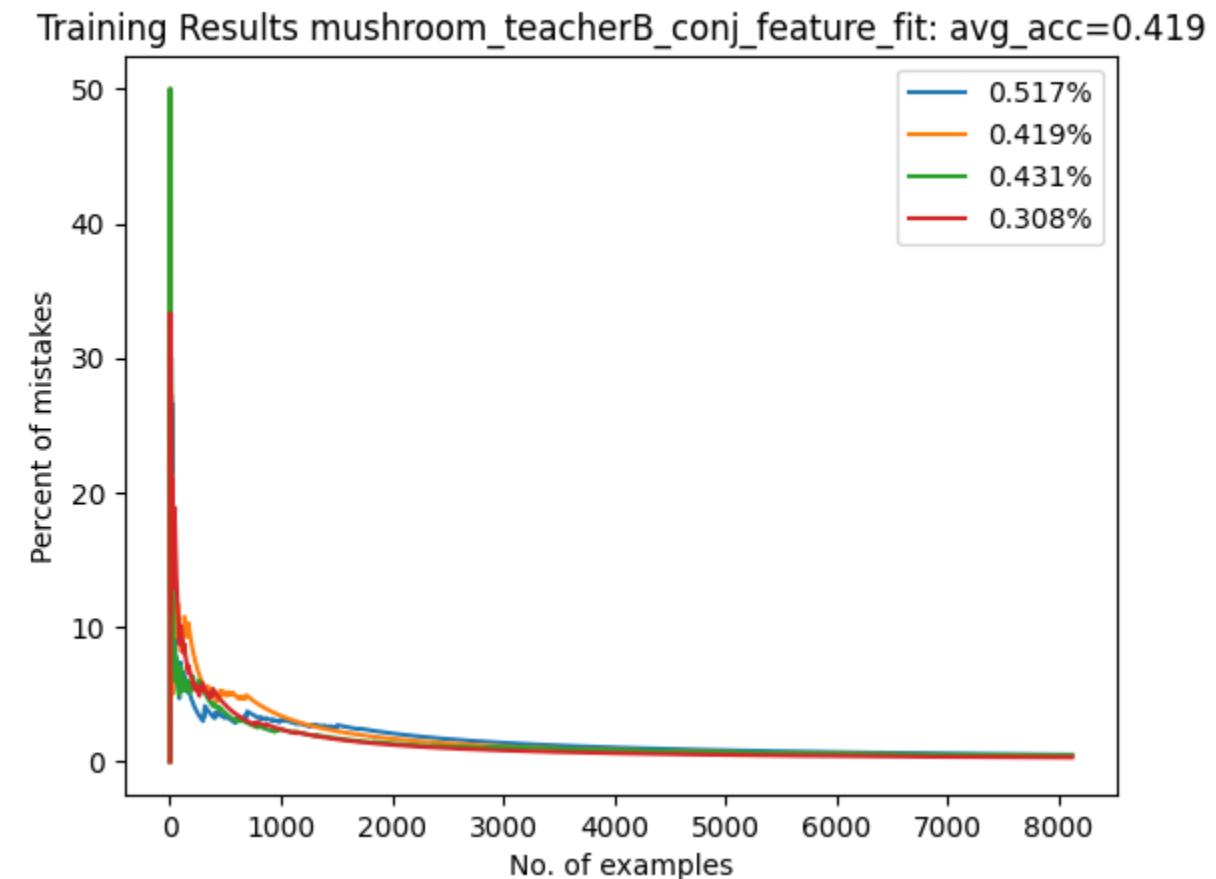
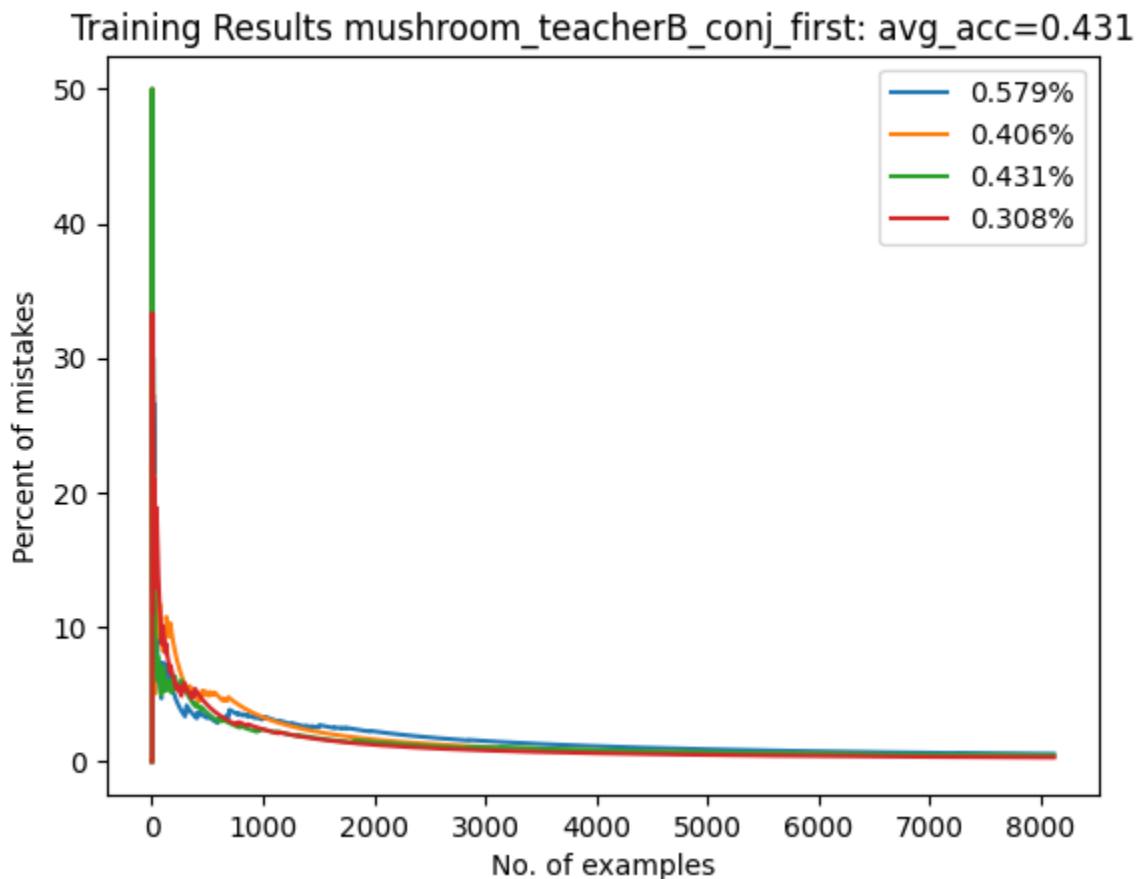
The Methods

- 1) choose a random conjunction from the conjunctions that fit the instance (class `ConjunctionFinderRandom`)
- 2) calculate the normalization of the fitting conjunctions and choose the one with the smallest absolute difference between the conjunction and the instance (class `ConjunctionFinderFeatureFit`) –
Motivation: if one of those conjunctions has the right label for the instance, it will probably be the one with the closest normalized features.
- 3) choose the conjunction with the biggest number of predicates. (class `ConjunctionFinderPredicatFit`) –
Motivation: the conjunction with the most predicates is the one with the most discriminative features because those features were picked by the Teacher, and the teacher always returns the most discriminative feature.
- all these options have a price in performance compared to find the first conjunction execution time

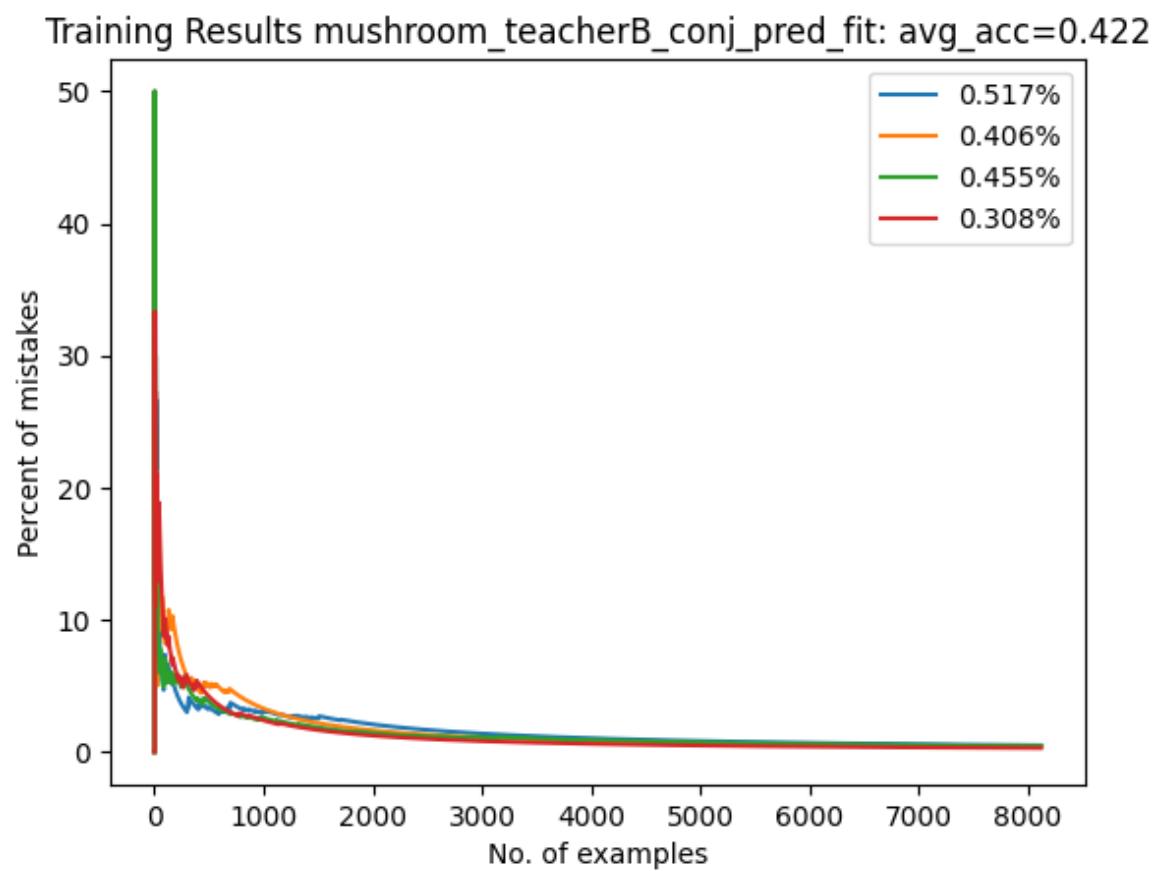
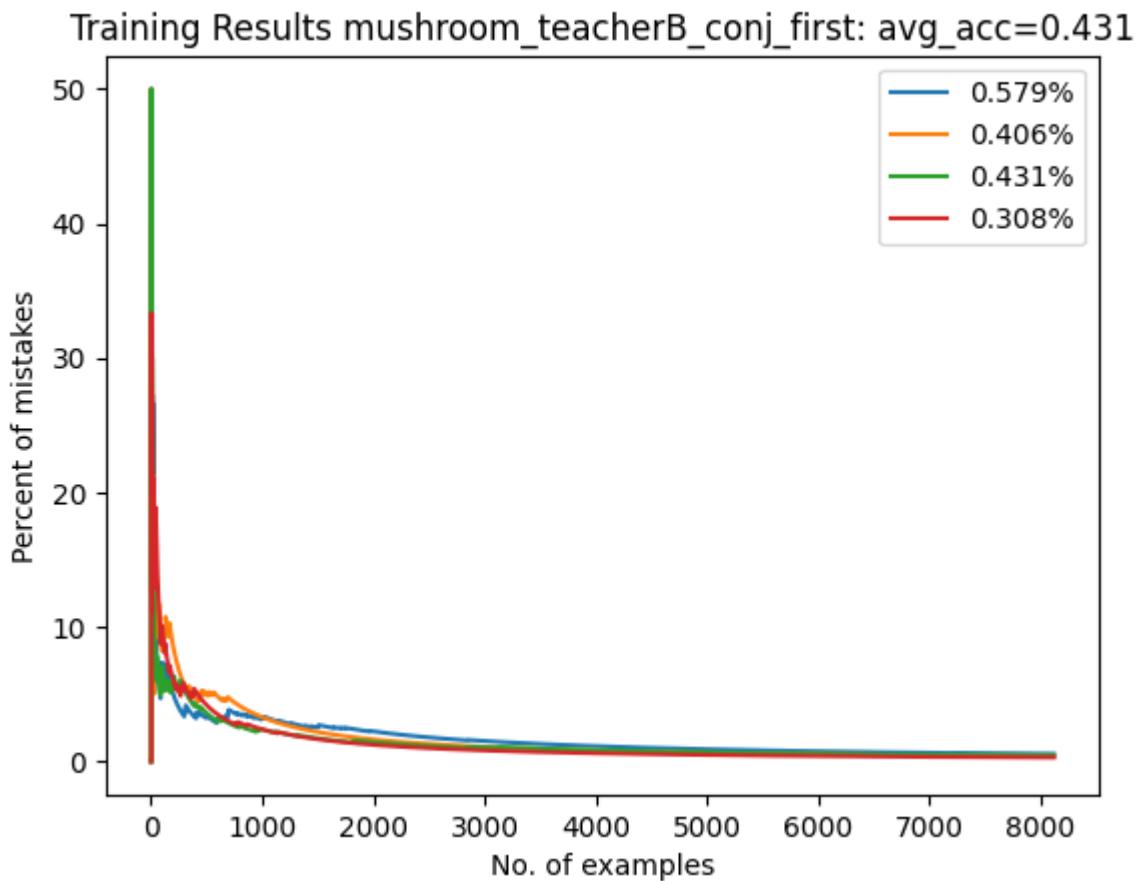
- Results of method1



- Results of method2



- Results of method3



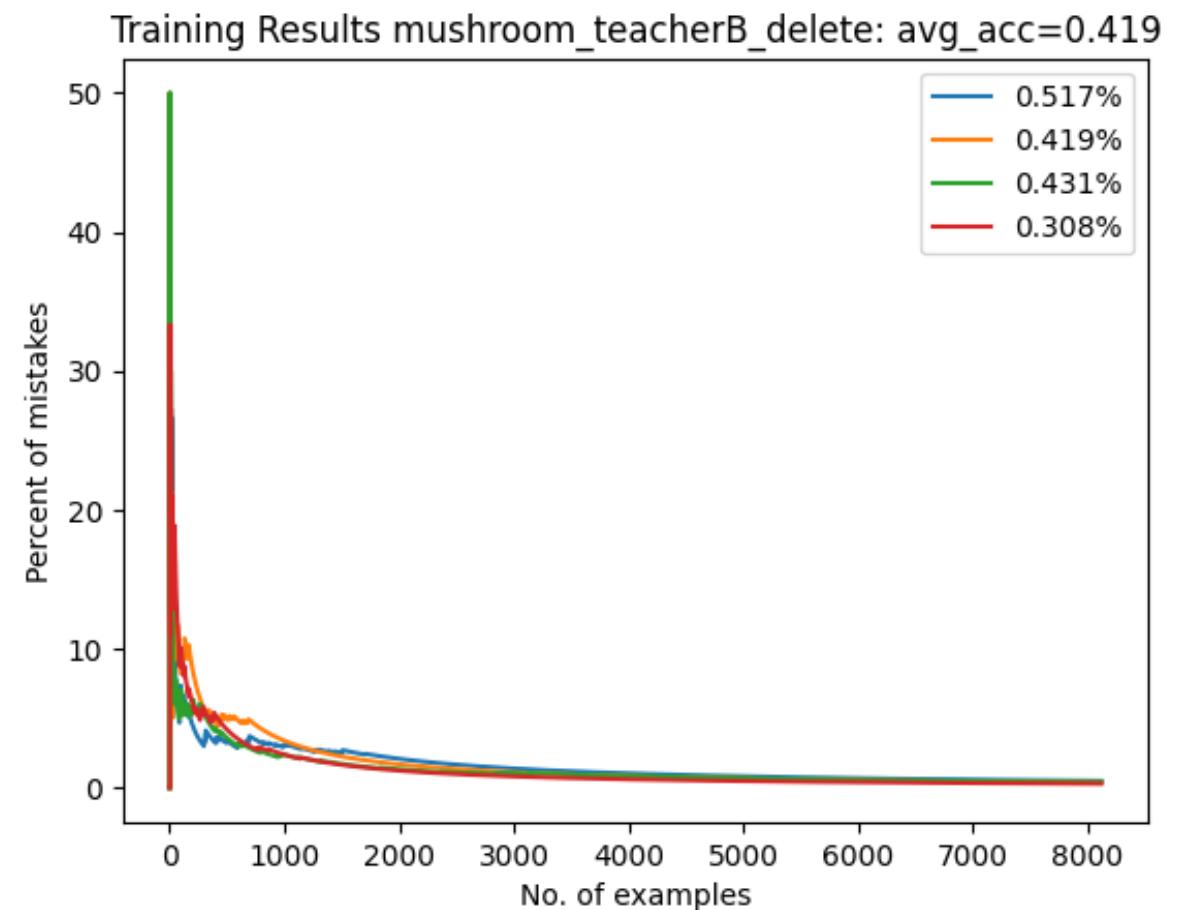
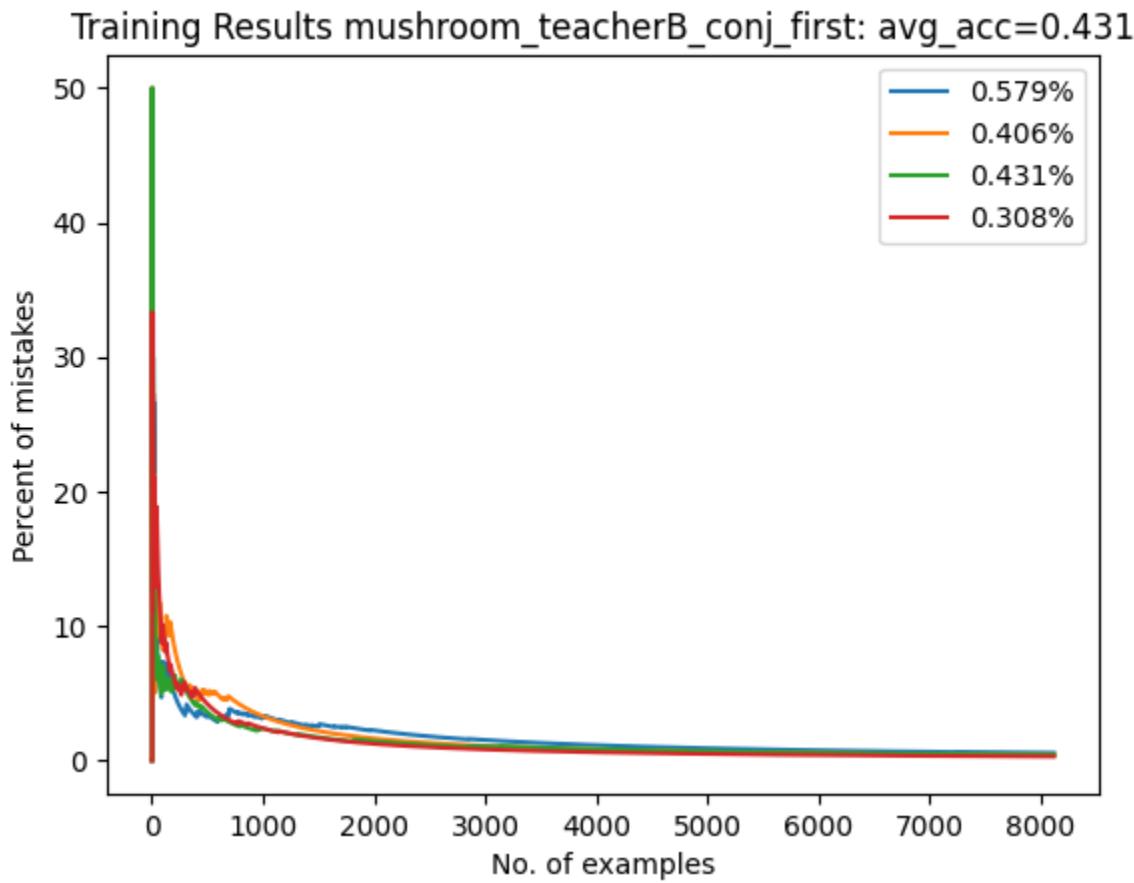
Conclusion

- On the Mushroom dataset Method1 was incorrect on average 0.437% of the times.
- On the Mushroom dataset Method2 was incorrect on average 0.419% of the times.
- On the Mushroom dataset Method3 was incorrect on average 0.422% of the times.
- The methods slightly improved TeacherB's accuracy but not in a very meaningful way.
- Our different ways of choosing the conjunction does not have much effect with the Mushroom dataset.
- Overall, Method2 was the most successful but not by a considerable margin.

3) quality of life control

- When we keep adding predicates to the conjunctions, we make it too specific for the instance that represents it and won't represent a subgroup of his label.
- Solution: deleting predicates.
- The learning algorithm receives a maximum misses number for conjunction and max number of predicates. We implement this when we guess a label for an instance and get it wrong, we find all the conjunctions with the same label as the current instance.
- We choose the one with the smallest relative number of predicates that don't fit the instance.
- In that conjunction, we find the predicates that caused the algorithm to not choose it and for each of them increase the misses counter by one.
- If there is a predicate that passes the max number of misses, we delete it
- If there is a conjunction that passes the max number of predicates, we delete the predicate with the most misses.

Results



Conclusion

- On the Mushroom dataset, the deletion method was incorrect on average 0.419% of the times.
- Deleting predicates improved the accuracy slightly.
- So, deleting predicates from too specific conjunctions improved the accuracy of the algorithm.

final summery

- Improving the teacher algorithm seems to have much better potential to improve accuracy then improving the learner algorithm.
- There are datasets which seems very suitable for these kind of algorithms, and some datasets which seems less suitable. In some cases, there is the option to change the dataset by adding more features to the dataset from exiting features combination (like smelly in mushrooms dataset).
- A more complicated dataset can probably get a good accuracy improvement by using a more sophisticated classifier in the teacher