

```

1 import numpy as np
2 import util
3 import sys
4
5 sys.path.append('../linearclass')
6
7 ### NOTE : You need to complete logreg implementation first!
8
9 from logreg import LogisticRegression
10
11 # Character to replace with sub-problem letter in plot_path/save_path
12 WILDCARD = 'X'
13
14
15 def main(train_path, valid_path, test_path, save_path):
16     """Problem 2: Logistic regression for incomplete, positive-only labels.
17
18     Run under the following conditions:
19     1. on t-labels,
20     2. on y-labels,
21     3. on y-labels with correction factor alpha.
22
23     Args:
24     train_path: Path to CSV file containing training set.
25     valid_path: Path to CSV file containing validation set.
26     test_path: Path to CSV file containing test set.
27     save_path: Path to save predictions.
28     """
29     output_path_true = save_path.replace(WILDCARD, 'true')
30     output_path_naive = save_path.replace(WILDCARD, 'naive')
31     output_path_adjusted = save_path.replace(WILDCARD, 'adjusted')
32
33     # *** START CODE HERE ***
34     plot_path = save_path.replace('.txt', '.png')
35     plot_path_true = plot_path.replace(WILDCARD, 'true')
36     plot_path_naive = plot_path.replace(WILDCARD, 'naive')
37     plot_path_adjusted = plot_path.replace(WILDCARD, 'adjusted')
38
39     # Part (a): Train and test on true labels
40     # Make sure to save predicted probabilities to output_path_true using np.savetxt()
41
42     x_train, t_train = util.load_dataset(train_path, label_col='t',
43                                         add_intercept=True)
44
45     clf = LogisticRegression()
46     clf.fit(x_train, t_train)
47
48     x_test, t_test = util.load_dataset(test_path, label_col='t',
49                                       add_intercept=True)
50
51     p_test = clf.predict(x_test)
52     np.savetxt(output_path_true, p_test)
53     util.plot(x_test, t_test, clf.theta, plot_path_true)
54
55     # Part (b): Train on y-labels and test on true labels
56     # Make sure to save predicted probabilities to output_path_naive using np.savetxt()
57
58     x_train, y_train = util.load_dataset(train_path, label_col='y',
59                                         add_intercept=True)
60
61     clf = LogisticRegression()
62     clf.fit(x_train, y_train)
63
64     x_test, t_test = util.load_dataset(test_path, label_col='t',
65                                       add_intercept=True)
66
67     p_test = clf.predict(x_test)
68     np.savetxt(output_path_naive, p_test)
69     util.plot(x_test, t_test, clf.theta, plot_path_naive)
70
71     # Part (f): Apply correction factor using validation set and test on true labels
72
73     x_valid, y_valid = util.load_dataset(valid_path, label_col='y')
74     x_valid = x_valid[y_valid == 1, :] # Restrict to just the labeled examples
75     x_valid = util.add_intercept(x_valid)
76     y_pred = clf.predict(x_valid)
77     alpha = np.mean(y_pred)

```

```
73 print('Found alpha = {}'.format(alpha))
74 x_test, t_test = util.load_dataset(test_path, label_col='t',
75                                   add_intercept=True)
76
77 # Plot and use np.savetxt to save outputs to output_path_adjusted
78 np.savetxt(output_path_adjusted, p_test / alpha)
79 util.plot(x_test, t_test, clf.theta, plot_path_adjusted, correction=alpha)
80
81 # *** END CODER HERE
82
83 if __name__ == '__main__':
84     main(train_path='train.csv',
85         valid_path='valid.csv',
86         test_path='test.csv',
87         save_path='posonly_X_pred.txt')
```