

# CSE 4392 Special Topic: Natural Language Processing

## Homework 8 - Spring 2025

March 26, 2025

### Problem 1 - 30%

Show that your HMM code from the last assignment is a maximum likelihood estimate by computing the likelihood of the training set over a large set of randomly selected parameter matrices and comparing its value with that obtained using the trained HMM parameters.

### Problem 2 - 70%

We have two biased coins:

- Coin A has a probability  $p$  of landing heads.
- Coin B has a probability  $q$  of landing heads.

In each experiment:

- A coin is randomly selected based on a mixing parameter  $\lambda$ :
  - With probability  $\lambda$ , Coin A is selected.
  - With probability  $1 - \lambda$ , Coin B is selected.
- The selected coin is tossed  $m$  times, and the number of heads is recorded.

The goal is to use the Expectation-Maximization (EM) algorithm to estimate the true parameters  $\lambda, p, q$  from the observed data (number of heads per experiment).

### Starter Code

```
import numpy as np

# -----
# Simulation settings:
# -----
np.random.seed(42)

# True parameters
```

```
true_lambda, true_p, true_q = 0.6, 0.7, 0.4
m, N = 10, 500 # m = tosses per experiment, N = number of experiments

# Simulate complete data
coin_choices = np.random.rand(N) < true_lambda
data = np.random.binomial(m, np.where(coin_choices, true_p, true_q))

# Complete assignments (1 for coin A, 0 for coin B)
assignments = coin_choices.astype(int)

# -----
# EM algorithm initialization:
# -----
lambda_est, p_est, q_est = 0.5, 0.6, 0.5
max_iters, tol = 200, 1e-6

# TODO: Implement the EM algorithm
# Ensure it converges to the right answer for at least one initialization
# and make a table of at least 10 rows detailing the number of iterations
# required for each of different initializations.
```

Preferably, submit a PDF notebook using this tool. Or submit a Python file and screenshot for the output. If your code output does not match the reported output, zero will be granted.