Small Problem 3: Discrete-time Discrete-observation HMM

Queries:

Query 1: The MAP state sequence.

Query 2: Smoothing: For each time step $t > 1$, the marginal distribution $P(S_t|O)$, where O is the output sequence.

Query 3: Filtering: For each time step $t > 1$, compute the marginal distribution $P(S_t|O_{1:t})$, where $O_{1:t}$ is the vector of outputs from time 1 up to the current time t.

Metrics:

Metric 1 (for query 1 only):
Minimum Hamming distance between the predicted and the given MAP state sequences. Notice that there are 18 true MAP state sequences provided in this solution.

Metric 2 (for queries 2 and 3):

a) Total variation distance between the true and computed marginal posteriors at each time step.
b) The mean and variance of the per-step total variation distance computed across all time steps.

Ground Truth:

The answers to the queries are in the attached spreadsheet (`problem-3-solution.xlsx`). The Excel format file is represented by three CSV files, one for each query, i.e. map, smoothing, and filtering.

TODO:

Compute Metric 1 for Query 1.
Compute Metrics 2a and 2b for Queries 2 and 3. We have provided evaluation code that computes Metrics 2a and 2b in Matlab and Java. Both the Matlab program and the Java JAR file work the same.

The provided Matlab evaluation code uses the following command line:

```
matlab -nospash -nojvm -nodisplay -nodesktop -r
"TVDScoreHMM(<number-of-states>, <number-of-timesteps>,
<ground-truth-csv-path>,<input-csv-path>,<output-csv-path>)"
```

The output will be a CSV file written to `<output-path>`.

Note that the Matlab program for Metric 1 is compatible with GNU Octave.

The provided JAR file uses the following command line:

```
java -jar TVDScoreHMM.jar <ground-truth-csv-path> <input-csv-path> <output-
csv-path>
```

Ground truth CSV and the input CSV should not have any row/column header; should only have 20 (number of time steps) rows and 5 (number of state) columns.

Submission:

The metric value should be computed for each elapsed time step (by calling the provided code or by implementing yourself). The metric value should be reported for several elapsed time steps. The number of elapsed time steps should be sufficient to establish an "informative profile".

For further details regarding submission of the metric and your code, please refer to the main CP4 problem description document, e.g. PPAML-Challenge-Problem-4.pdf.

Sample output files for this problem have been provided in the "sampleoutput" folder:

```
problem-3-query-1-metric-1.csv
problem-3-query-2-metric-2.csv
problem-3-query-3-metric-3.csv
```

Submit the metric and your code as described in the main CP4 problem description document, e.g. PPAML-Challenge-Problem-4.pdf.

Ground Truth Details:

This discrete-time Discrete-observation HMM problem was solved using Kevin Murphy's Bayes Net Toolbox (BNT) for MATLAB (Murphy, 2001).