Assignment 6

Computational Intelligence, SS2017

Team Members		
Last name	First name	Matriculation Number
Reeh	Lucas	00630182

Contents

List of Figures

1 Viterbi Algorithm and optimal State Sequence

2. Optimal state sequences[1][2]

   Result for HMM_1

   ```
   X1 = [0, 0, 1, 1, 1, 0]
   q1 = ['r', 'r', 's', 's', 's', 's']
   X2 = [0, 0, 1, 1, 1, 0, 0]
   q2 = ['r', 'r', 'f', 'f', 'f', 'r', 'r']
   ```

   Result for HMM_2

   ```
   X1 = [0, 0, 1, 1, 1, 0]
   q1 = ['r', 'r', 'f', 'f', 'f', 'r']
   X2 = [0, 0, 1, 1, 1, 0, 0]
   q2 = ['r', 'r', 'f', 'f', 'f', 'r', 'r']
   ```

   Findings

   When observing X_1 in HMM_2 higher emission probaility for "no umbrella" on foggy days (0.95) and higher transition probaility for switching from rain to fog (0.6) leads to a different path through the states ($n-1$ to n) after switching observation from 2^{nd} day with umbrella to the 3^{rd} wihtout. Also the probability to stay in state sunny in HMM_1 when observing "no umbrella" is quite high.

   Optimal state sequences for observation X_2 seems to be no different for both models.

3. Numerical instability

   Since probability values in each iteration are getting smaller at a certain point numerical errors would appear. But if we are only interested in $argmax$ and max this problem can be avoided by using log for transition and emission probabilities without changing our results [3, p. 373] (104, 105a-c). This scaling method can be further improve as you can see in "Numerically Stable Hidden Markov Model Implementation"[1] to fix even more problems with log.

[1] 0 = umbrella, 1 = no umbrella
[2] r = rainy, s = summy, f = foggy

2 Sequence Classification

1. $P(X|\mathbf{\Theta_i})$

$$P(X|\mathbf{\Theta_i}) = \sum_{\mathbf{Q} \in \mathcal{Q}} P(X, \mathbf{Q}|\mathbf{\Theta_i})$$

With Q being the set of all state sequences. This would be very expensive to compute because of the exponential amount of sequences $|\mathcal{Q}| = (N_s)^N$. With naive marginalisation this would result in $\mathcal{O}(2N(N_s)^N)$ operations[2].

2. Joint probability

$$\begin{aligned} P(X|\mathbf{\Theta_i}) &= \sum_{\mathbf{Q} \in \mathcal{Q}} P(X, \mathbf{Q}|\mathbf{\Theta_i}) \\ &= \sum_{\mathbf{Q} \in \mathcal{Q}} P(X|\mathbf{Q}, \mathbf{\Theta_i}) P(\mathbf{Q}|\mathbf{\Theta_i}) \end{aligned}$$

So viterbi-algorithm already holds an approximation for $\mathcal{Q}_N$ after computing δ_N (max probability for partial sequences). The last observation can be used for classification when assuming the probability along the best path.

3. Classification

Result for X_1

$$\begin{aligned} P(X_1|\text{HMM}_1) &= 0.0020198 \\ P(X_1|\text{HMM}_2) &= 0.0008406 \\ &\implies \text{best is HMM}_1 \end{aligned}$$

Result for X_2

$$\begin{aligned} P(X_2|\text{HMM}_1) &= 0.0004099 \\ P(X_2|\text{HMM}_2) &= 0.0004062 \\ &\implies \text{best is HMM}_1 \quad \text{(seems wrong?)} \end{aligned}$$

3 Samples from a Gaussian Mixture Model

1. Examples for Sampling[3][4]

   ```
   Sampling from HMM1
   ['nu', 'u', 'u', 'nu', 'nu']
   ['r', 'r', 'r', 's', 's']
   Sampling from HMM2
   ['nu', 'u', 'nu', 'nu', 'u']
   ['f', 'r', 'f', 'f', 'r']
   ```

2. Same procedure as used as for GMM sampling: Piped usage of discrete sampling (with provided function). First draw for emission probability and transition probability and then draw depending on that every N-th sample (in GMM it was first distribution and depending on this the second and so on ...).

[3] u = umbrella, nu = no umbrella

[4] r = rainy, s = summy, f = foggy

4 Markov Model

1. Compute P_2 and P_3 (Markov-Chain)

$$P_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$
$$P_2 = \pi_1 \cdot A$$
$$P_2 = \begin{bmatrix} 0.4 \\ 0.31\dot{6} \\ 0.28\dot{3} \end{bmatrix}$$
$$P_3 = \pi_1 \cdot A^2$$
$$P_3 = \begin{bmatrix} 0.24 \\ 0.1508\dot{3} \\ 0.10416\dot{6} \end{bmatrix}$$

2. P_n only works for countable infitie state transitions (Markov-Chain)

$$P_{n+1} = \pi_1 \cdot A^{n-1}$$

or if you shift the index

$$P_0 = \pi_0$$
$$P_n = \pi_0 \cdot A^n$$

References

[1] Tobias P. Mann. “Numerically stable hidden Markov model implementation”. In: *Ms. Feb* (2006).

[2] Franz Pernkopf. *Computational Intelligence: Teil 2 (Vorlesungsmitschrift)*. [Online; Stand 29. Juni 2016]. 2016.

[3] Lawrence R. Rabiner. “Readings in Speech Recognition”. In: ed. by Alex Waibel and Kai-Fu Lee. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990. Chap. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, pp. 267–296. ISBN: 1-55860-124-4. URL: `http://dl.acm.org/citation.cfm?id=108235.108253`.