# DD2380 grading sheet: HMM

PLEASE USE BLOCK LETTERS

| | | | | | |
|---|---|---|---|---|---|
| TA | | | Date & time | | |
| Reported in Canvas: | Student 1 | | | Student 2 | |

| | Student 1 | Student 2 |
|---|---|---|
| Name | | |
| KTH id (email name) | | |
| Personal identity number | | |
| Kattis submission id (hmm3/Fishing derby) | | |
| Aiming for grade | | |
| Grade | | |

| | |
|---|---|
| Contributed equally | |

| | | Student 1 | Student 2 |
|---|---|---|---|
| | Brought written solutions to questions 1-6 from the assignment description | | |
| | HMM0 – HMM3 OK in Kattis | | |
| | Wrote (parts of) the code | | |
| E | Can explain the code | | |
| D | Can answer basic questions about HMMs | | |
| | Brought written solutions to questions 7-10 from the assignment description | | |
| | Can justify all design decisions about the code | | |
| C | Can answer more involved questions about HMMs based on empirical investigations | | |
| | Fishing derby > 250 in Kattis | | |
| B | Can answer advanced questions about HMMs | | |
| | Fishing derby > 250 in Kattis by **deadline** | | |
| A | Can suggest how to address modified versions of Fishing derby | | |

| Answers: | Student 1 | | | | | Student 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Fully | With a little help | Partially or with a lot of help | Not | | Fully | With a little help | Partially or with a lot of help | Not |

| Questions basic theory | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| **Questions more involved** | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| **Questions advanced** | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

| Notes | |
|---|---|
| Student 1 | |
| Student 2 | |

## **Guidelines**:

**General**:

- Always fill out the protocol
- Always conduct the presentation in English unless they specifically asked for Swedish (we are legally obliged to provide Swedish examination if required…)
- Target your questions to one student at a time, switch turns
- Try hard to stick to 20 minutes max. and recommended max. no. of questions

- After the presentation, put the grade in Canvas (ideally within 3 days) and bring the protocol to Iolanda (or slip into her mailbox at 3rd floor Teknikringen 33).

**How to proceed during the presentation:**

- Fill out the data and always check the names against their ID cards
- Start easy: “How are you doing, how did it go?”
- Ask what grade they are aiming for
- Ask how they split the work; if they did not contribute equally, if unusual make a note in the Notes section
- Ask about HMM code first.
- See if they brought solutions to questions from the assignment description
- Proceed with D-level questions, C-level, then move to Fishing derby and finish up with B-level and A-level questions; put the numbers of the questions from the bank in the table
- How many and which questions to answer: ca 3 questions per person; some are super-quick, some are more involved

| | | No of questions from bank per pair | | | |
|---|---|---|---|---|---|
| | | Basic theory | More involved | Advanced | |
| | | D-level | C-level | B-level | A-level |
| Aiming for grade | D | 4-6 | | | |
| | C | 1-2 | 2-4 | | |
| | B | 1 | 1 | 2-4 | |
| | A | 1 | 1 | 1 | 2-3 |

**Filling the tables in the protocol:**

- Grading answers to questions:
  - **Fully**: knows immediately how to answer, no hesitation, no speculation, answers confidently
  - **With a little help**: needed to reformulate the question, give some more details what is meant by the question, but knows the answer, answers fully
  - **Partially or with a lot of help**: almost more work for us than for them to answer, but we get the answer out somehow, at least partially
  - **Not**: extreme, no idea

**In the checklist:**

| | |
|---|---|
| Can answer basic questions about HMMs | If most of the D-level asked questions are answered at most "with a little help" and if none of the asked questions are left not answered |
| Can answer more involved questions about HMMs based on empirical investigation | If most of the C-level asked questions are answered at most "with a little help" and if none of the asked questions are left not answered |
| Can answer advanced questions about HMMs | If most of the B-level asked questions are answered at most "with a little help" and if none of the asked questions are left not answered |
| Can suggest how to address modified versions of Fishing derby | If most of the A-level asked questions are answered at most "with a little help" and if none of the asked questions are left not answered |

**Other less obvious things in the checklist:**

| | |
|---|---|
| Wrote (parts of) the code | Ask specifically which parts of code they wrote |
| Can explain the code | Ask to explain overall, but do not let them go on and on for ages, complement with quick questions like what does this procedure do, or where is a procedure that does this and that |
| Can justify all design decisions about the code | Ask questions, such as how did you initialize the parameters, what exactly is happening in this Viterbi algorithm |

**Final grade:**
E: checks everything in block 1 of the checklist
D: checks everything in block 1 + 2 of the checklist
C: checks everything in block 1 + 2 + 3 of the checklist
B: checks everything in block 1 + 2 + 3 + 4 of the checklist
A: checks everything in block 1 + 2 + 3 + 4 + 5 of the checklist

*Potential exception:* did not bring written solutions to the questions, but know the answers fully; in that case count it as check, but leave a note in the Notes section

**Notes section:**
- If you answer a question out of our question banks, put it here
- If you want to leave some special feedback, put it here and later on put as a comment to the grade in canvas
- If you think the person would be a great TA next year, leave a comment
- If the code was extra efficient, leave a comment.

# Question Bank

## D-level:

1. What property should the transition matrix and the emission matrix have?
2. What are the parameters of an HMM?
3. What is the Markov property?
4. What does the transition matrix of an HMM describe?
5. What does the emission matrix of an HMM describe?
6. What is the meaning of the initial state distribution?

## C-level:

1. How do you initialize the HMM parameters and why?
2. Why can't you initialize the matrices with a uniform distribution?
3. What would happen if you set A(i,i) = 1.0 for example?
4. Have you thought about convergence criteria? What does it mean here that the algorithm converges?
5. Does the algorithm always converge?
6. Does it always converge to the actual original HMM?

## B-level:

1. Do you reinitialize your model every time you get a new observation in the observation sequence?
   Here we accept two options:
   - Yes. Because we can learn the model that best fits the sequence O_1..t without any observation of O_i = k and that might end up with zeros in the model. That way, we might not be able to move from that bad local minima to find a better solution for our model.
   - No. We want to use the previous information. However, in order to be safe from a bad local minima we add some random noise (or perturbation) to the previous model solution.
2. Consider the original Fishing derby game, where the goal is to catch fish. There is a very rare fish species circulating in the water, the angel shark, which you must not catch. How would you implement this?
   Any reasonable strategy should work. E.g., use your guessing algorithm to compute the confidence the fish is not the angel shark (i.e., P(not angel shark | models, observations)) and only catch when this confidence is high.
3. How should you define convergence? When are you sure enough to guess? What was your strategy for guessing?
   - Convergence is defined again as when P(observations | model) stops increasing (or in a CS real scenario with floating point errors when the absolute difference between the previous and current likelihood is smaller than epsilon).
4. What happens if we decrease or increase the number of states?
   - If we decrease the number of states we end up with state collapse; i.e., some states capture information from more than one real state (e.g., in an application where states represent real values we might go from states x, y, z to states u = x + 2y, v = -z + y).
   - If we increase the number of states we end up with uninformative states and hence sparse matrices.
5. What priors do you have for the initializations of the A, B and pi?
   Again we accept two main solutions:
   - Uniform matrices with a small random noise perturbation to avoid a bad local minima.
   - Uniform matrices with a small random noise perturbation for B and pi and a diagonal matrix for A with a small random noise perturbation to avoid local minima. Matrix A initialized this way only if they thought about the hidden states as the swimming patterns and they want to use the prior knowledge that transitioning from one swimming pattern to another is unlikely.

6. Explain how you compute the most probable next emission.
   You compute the most likely next emission and the model given all the observations and then filter out for all the models. i.e., sum(P(O_next, model | observations)) for all models. Alternatively, take the maximum, but this second approach is worse.
7. Is there an alternative way to keeping a history of all the HMMs you've ever learned?
   Yes, you could only take one HMM per species and use the observations from each fish on that species to train it. For instance, stacking the observations observed for that species and training the model with that. This way you would end up with longer observations.

## A-level:

1. What if vertical and horizontal movements were conditionally independent? How could you exploit this?
   Then matrix B would be a sparse matrix. We could exploit this by having two different observation matrices (B_h and B_v) standing for horizontal and vertical observation matrices. These matrices would predict independently the movements left, right, stay and up, down, stay so that, for instance left + down = diagonal sud-west.
   This is an exploitation because the training and storage would be smaller since the two matrices are 2*(3*H) instead of 9*H with H the number of hidden states.
2. Do you consider comparing the model matrices for fish classification? Why can you not compare the matrices directly to identify the fish?
   Here many reasons might apply. For instance, we know that the hidden states do not have a defined order and the matrices A1, B1, pi1 are equivalent to the matrices A2, B2, pi2 when the hidden states are in different order (i.e., A2, B2, pi2 are obtained by the same row permutation from A1, B1, pi1).
3. Consider the sampling rate of the observations. How would A, B, and pi change in case of an extremely low or high sampling rate?
   **-** Sampling in a very high sampling rate might cause over-representation of some samples and might be difficult for the algorithm to predict the future based on the present since predicting a stationary observation would be very likely.
   - Sampling in a very low frequency rate might cause under-representation of some samples and hence the algorithm to miss important patterns in the observations.
   Lets take the following example:
   U L U R U L U R (pattern at a right sampling rate)
   UUUUULLLLLLUUUUURRRRRUUUULLLLLUUUUUUR (pattern at a high sampling rate)
   U U U U (pattern at a low sampling rate)
4. TA suggests a different problem that HMMs are frequently used for (e.g. hand writing, speech music, forecasting weather etc.). What are the states and emissions in this problem?
   Here just be creative and put an example you are comfortable with. Feel free to use the ones we used in the tutorials.
5. There is algae bloom in the water and some emissions can't be observed. What problem would you have to think about then?
   - When training, use the sequence of observations O_1..T to learn as usual. However, now the unobserved observations will have to be estimated using both the observations in the past and the future (i.e., the student should notice that P(X_t | X_1..t-1) = P(X_t | X_t-1) and P(X_t | X_t+1..T) = P(X_t | X_t+1). This should be done at each iteration of the algorithm with the new model.
   - For the other problems the unobserved observations only have to be estimated once in order to proceed with the algorithm.