# WASP SUMMER SCHOOL ON MACHINE LEARNING

# LAB: BNS AND HMMS

JOSE M. PEÑA
LINKÖPING UNIVERSITY, SWEDEN

## 1. INSTRUCTIONS

The lab is designed to be solved with the R package `HMM`. Students may also want to use the RStudio development environment.

- Literature:
  - Package documentation.
- Hint: Spend the first 30 minutes reading the package documentation at `https://cran.r-project.org/web/packages/HMM/HMM.pdf`. You may even want to try the dishonest casino example included in the package.

## 2. QUESTIONS

The purpose of the lab is to put in practice some of the concepts covered in the lectures. To do so, you are asked to model the behavior of a robot that walks around a ring. The ring is divided into 10 sectors. At any given time point, the robot is in one of the sectors and decides with equal probability to stay in that sector or move to the next sector. You do not have direct observation of the robot. However, the robot is equipped with a tracking device that you can access. The device is not very accurate though: If the robot is in the sector $i$, then the device will report that the robot is in the sectors $[i-2, i+2]$ with equal probability.

(1) Build a HMM for the scenario described above.
(2) Simulate the HMM for 100 time steps.
(3) Discard the hidden states from the sample obtained above. Use the remaining observations to compute the filtered and smoothed probability distributions for each of the 100 time points. Compute also the most probable path.
(4) Compute the accuracy of the filtered and smoothed probability distributions, and of the most probable path. That is, compute the percentage of the true hidden states that are guessed by each method.
    Hint: Note that the function forward in the `HMM` package returns probabilities in log scale. You may need to use the functions `exp` and `prop.table` in order to obtain a normalized probability distribution. You may also want to use the functions `apply` and `which.max` to find out the most probable states. Finally, recall that you can compare two vectors A and B elementwise as `A==B`, and that the function `table` will count the number of times that the different elements in a vector occur in the vector.
(5) Repeat the previous exercise with different simulated samples. In general, the smoothed distributions should be more accurate than the filtered distributions. Why ? In general, the smoothed distributions should be more accurate than the most probable path, too. Why ?
(6) Is it true that the more observations you have the better you know where the robot is ?
    Hint: You may want to compute the entropy of the filtered distributions with the function `entropy.empirical` of the package `entropy`.
(7) Consider any of the samples above of length 100. Compute the probabilities of the hidden states for the time step 101.