Carnegie Mellon University, Machine Learning 10-701 — April 13, 2015
http://alex.smola.org/teaching/10-701-15 — due on April 20, 2015

Homework 11

START HERE: Instructions

- The homework is due at 9:00am on April 20, 2015. Anything that is received after that time will not be considered.
- Answers to every theory questions will be also submitted electronically on Autolab (PDF: Latex or handwritten and scanned). Make sure you prepare the answers to each question separately.
- Collaboration on solving the homework is allowed (after you have thought about the problems on your own). However, when you do collaborate, you should list your collaborators! You might also have gotten some inspiration from resources (books or online etc...). This might be OK only after you have tried to solve the problem, and couldn't. In such a case, you should cite your resources.
- If you do collaborate with someone or use a book or website, you are expected to write up your solution independently. That is, close the book and all of your notes before starting to write up your solution.

1 Hidden Markov Model (HMM) v.s. Linear Gaussian State Space Models (SSM) [70 pts]

We introduce SSM which share the same topology as HMMs. But the difference is that in HMM the states are discrete (multinomial), while in SSM the states follow continuous (Gaussian) distributions. We use the following notations:

- N - the number of states in the model
- X - the state sequence, $X = \{X_1, X_2, ..., X_N\}$. We
- V - the vocabulary of size M, $V = \{V_1, V_2, ...V_k, ..., V_M\}$
- A - the state transition matrix, $A = \{a_{ij}\}$
- Y - the observation sequence of length T, $Y = \{Y_1, Y_2, ..., Y_T\}$

For HMM,

- B - the emission matrix, $B = \{b_{jk}\}$, b_{jk} is the probability of emitting V_k when in state j.

For SSM,

- G - the transition error matrix, such that $X_t = AX_{t-1} + GW_t$, where $W_t = N(0, Q)$ is some Gaussian noise.
- C - the loading matrix, such that $y_t = Cx_t + v_t$, where $v_t = N(0, R)$ is some Gaussian noise.
- d - the dimension of the hidden state of SSM

Notice that we make no assumptions on Q or R, and also we assume $x_0 = N(0, \Sigma_0)$

We are interested in computing $p(x_t|y_{1:t})$, which corresponds to the second problem in HMM, namely the decoding problem. But notice that since we are computing $p(x_t|y_{1:t})$ instead of $p(x_t|y_{1:T})$, therefore it is a filtering (real-time estimation) problem rather than a smoothing problem, which means the backward variable is not needed.

1.1 Forward Inference for HMM [5+10=15 pts]

Please derive the forward variable (α) used in the decoding step in HMM, and use it to infer $p(x_t|y_{1:t})$.

1.2 Kalman Filtering for SSM [5+10+30+10=55 pts]

- Please briefly formulate a SSM with the provided notations (you may introduce new variables if necessary)
- Please give an intuitive and high-level description of the Kalman Filtering algorithm.
- Please infer $p(x_t|y_{1:t})$ with Kalman Filtering for SSM. Notice that $p(x_t|y_{1:t})$ follows a Gaussian distribution because everything in a SSM is assumed to be Gaussian. Therefore this reduces to finding the mean and covariance of this distribution.
  (You are required to derive the *prediction step* and the *update step*, and briefly address how the recursion works.)
- Please discuss the complexity of the Kalman Filtering algorithm, and answer in big-O notations.

2 Collaborative Filtering (CF) [30 pts]

2.1 Cold Start Problem [10 pts]

Please briefly describe what a cold start problem is in the context of model-based CF (matrix factorization).

2.2 Strategies [10+10=20 pts]

Please propose two different strategies to address the cold start problem.