CS 7641 CSE/ISYE 6740 Homework 4

Rakesh Surapaneni

Deadline: 11/29 Sunday, 11:55 pm

- Submit your answers as an electronic copy on T-square.
- No unapproved extension of deadline is allowed. Late submission will lead to 0 credit.
- Typing with Latex is highly recommended. Typing with MS Word is also okay. If you handwrite, try to be clear as much as possible. No credit may be given to unreadable handwriting.
- Explicitly mention your collaborators if any. For the programming problem, it is absolutely not allowed to share your source code with anyone in the class as well as to use code from the Internet without reference.
- Recommended reading: PRML Section 13.2

1 Kernels [20 points]

This problem will explore a number of kernels and non-kernels to get some more intuition for (i) what constitutes a valid kernel and (ii) see kind of functions we can implicitly define with kernels. Identify which of the followings is a valid kernel. If it is a kernel, please write your answer explicitly as 'True' and give mathematical proofs. If it is not a kernel, please write your answer explicitly as 'False' and give explanations.
Suppose K_1 and K_2 are valid kernels (symmetric and positive definite) defined on $R^m \times R^m$.

1. $K(u, v) = \alpha K_1(u, v) + \beta K_2(u, v), \alpha, \beta \in R$.
   Answer:
   False, It is not a valid Kernel.
   Let the gram matrix (Pairwise kernel) matrix for K be G, K_1 be G_1 and K_2 be G_2.
   We know that $K(u, v) = \alpha K_1(u, v) + \beta K_2(u, v), \alpha, \beta \in R$
   Clearly

   $$G = \alpha G_1 + \beta G_2$$

   $$=> v^T G v = \alpha v^T G_1 v + \beta v^T G_2 v$$

   If both α and β are negative, then $v^T G v$ will be negative since $v^T G_1 v, v^T G_2 v$ are positive values since they are Gram matrix for valid kernels.
   Hence unless both α and β are non negative, K need not be a valid kernel.

2. $K(u,v) = u^\top Cv$ where $C \in R^{m\times m}$. **Answer:**
False, It is not a valid Kernel.
We know that $u^\top v$ is a valid kernel. Now if C = -1*I, then $K(u,v) = -u^\top v$, which is not a valid kernel by same logic as in problem 1. ($\alpha = -1$ and $\beta = 0$). Hence the above kernel is valid only for some specific cases of C.

3. $K(u,v) = K_1(f(u), f(v))$ where $f : R^m \to R^m$. **Answer:**
True, It is a valid Kernel.
Note that K is a kernel since K_1 is a valid kernels and $f(u), f(v)$ are valid points in same space R^m.
In other words, let $K_1(u,v) = \phi(u)^\top \phi(v)$, then $K(u,v) = \phi_f(u)^\top \phi_f(v)$, where $\phi_f(u) = \phi(f(u))$.
Hence K is a valid kernel.

4. $K(u,v) = f(K_1(u,v))$ where f is any polynomial with positive coefficients. **Answer:**
True, It is a valid Kernel.
To prove this, **first** we prove that $(K_1(u,v) * k_2(u,v))$, i.e, product of two kernel functions, is kernel.
If we expand the definitions of both kernels, we get $K_1(u,v) = \sum_j \phi_1^j(u) * \phi_1^j(v)$ and $K_2(u,v) = \sum_i \phi_2^i(u) * \phi_2^i(v)$. The product is $K_1(u,v) * K_2(u,v) = \sum_{i,j} \phi_1^j(u)\phi_2^i(u) * \phi_1^j(v)\phi_2^i(v)$ This is a valid kernel with feature function $\phi_{ij}(u) = \phi_1^j(u)\phi_2^i(u)$ has m*n dimensions, and where m,n are dimensions of ϕ_1 and ϕ_2 respectively.
Secondly, from part a, we know that weighted sum of various kernel functions is a kernel, if weights are positive.
Hence we can prove that, $K(u,v) = f(K_1(u,v))$ where f is any polynomial with positive coefficients, **is a valid kernel.**

5. $K(u,v) = \exp K_1(u,v)$.
Answer:
True, It is a valid Kernel.

We can write an expression of limit equation for exponential as a limit to polynomial (Taylor series)

$$\exp x = \lim_{i\to\infty} (1 + \frac{x^2}{2!} .... + \frac{x^i}{i!})$$

Using result from above, we can prove that gram matrix for exponential is non negative and hence exponential is also a valid kernel function.

6.

$$K(u,v) = \begin{cases} 1 & \text{if } \|u-v\|_2 \leqslant 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Answer:
True, It is a valid Kernel.

The Gram matrix for this function is similar to the gram matrix for kernel$u^T v$, where each point is either (1) or (0), Definitely, since the $v^T G v$ is always non negative for second gram matrix, so is true for the first one.
Hence we have symmetric non negative semi-definite Gram matrix which makes our function a valid kernel.

7. Suppose K' is a valid kernel.

$$K(u, v) = \frac{K'(u, v)}{\sqrt{K'(u, u)K'(v, v)}}. \tag{2}$$

Answer:
True, It is a valid Kernel.

Let $K'(u, v) = \phi^T(u)\phi(v)$, then $K(u, v) = \phi_1^T(u)\phi_1(v)$ where $\phi_1 = \phi(u)/\sqrt{K'(u, u)}$. Hence K is a valid kernel function.

2 Markov Random Field, Conditional Random Field [20 pts]

[a-b] A probability distribution on 3 discrete variables a,b,c is defined by $P(a,b,c) = \frac{1}{Z}\psi(a,b,c) = \frac{1}{Z}\phi_1(a,b)\phi_2(b,c)$, where the table for the two factors are given below.

a	b	$\phi_1(a,b)$
0	0	4
0	1	3
1	0	3
1	1	1

b	c	$\phi_2(b,c)$
0	0	3
0	1	2
0	2	1
1	0	4
1	1	1
1	2	3

(a) Compute the slice of the joint factor $\psi(a,b,c)$ corresponding to $b = 1$. This is the table $\psi(a, b = 1, c)$. [5 pts]

Solution:

$$\psi(a,b,c) = \phi_1(a,b) * \phi_2(b,c)$$

Hence

$$\psi(a, b = 1, c) = \phi_1(a, b = 1) * \phi_2(b = 1, c)$$

Using the above equation, there are 6 possible values we need to compute, (2 values for a and 3 values for c). Therefore the table is

a	c	$\psi(a, b = 1, c)$
0	0	3*4 = 12
0	1	3*1 = 3
0	2	3*3 = 9
1	0	1*4 = 4
1	1	1*1 = 1
1	2	1*3 = 3

(b) Compute $P(a = 1, b = 1)$. [5 pts]

Solution:
First we calculate Z,

We know that,

$$P(a,b,c) = \frac{1}{Z}\psi(a,b,c) = \frac{1}{Z}\phi_1(a,b)\phi_2(b,c)$$

$$=> Z = \sum_{ain\{0,1\},b\in\{0,1\},c\in\{0,1,2\}} \phi_1(a,b)\phi_2(b,c)$$

$$= \sum_{ain\{0,1\},c\in\{0,1,2\}} \phi_1(a, b = 1)\phi_2(b = 1, c) + \sum_{ain\{0,1\},c\in\{0,1,2\}} \phi_1(a, b = 0)\phi_2(b = 0, c)$$

$$= \sum_{ain\{0,1\}} \phi_1(a, b = 1) * \sum_{c\in\{0,1,2\}} \phi_2(b = 1, c) + \sum_{ain\{0,1\}} \phi_1(a, b = 0) * \sum_{c\in\{0,1,2\}} \phi_2(b = 0, c)$$

$$\sum_{ain\{0,1\}} \phi_1(a, b = 1) = 3 + 1 = 4$$

$$\sum_{ain\{0,1\}} \phi_1(a, b = 0) = 4 + 3 = 7$$

$$\sum_{c\in\{0,1,2\}} \phi_2(b = 1, c) = 4 + 1 + 3 = 8$$

$$\sum_{c\in\{0,1,2\}} \phi_2(b = 0, c) = 3 + 2 + 1 = 6$$

Therefore

$$Z = \sum_{ain\{0,1\}} \phi_1(a, b = 1) * \sum_{c\in\{0,1,2\}} \phi_2(b = 1, c) + \sum_{ain\{0,1\}} \phi_1(a, b = 0) * \sum_{c\in\{0,1,2\}} \phi_2(b = 0, c) = 4*8+7*6 = 32+42 = 74$$

$$P(a = 1, b = 1) = \sum_{c\in\{0,1,2\}} P(a = 1, b = 1, c) = 1/74 * \sum_{c\in\{0,1,2\}} \phi_2(a = 1, b = 1)\phi_2(b = 1, c) = 8/74 = 4/37$$

$\boxed{P(a = 1, b = 1) = 4/37 = 0.108}$

(c) Explain the difference between Conditional Random Fields and Hidden Markov Models with respect to the following factors. Please give only a one-line explanation. [10 pts]

- Type of model - generative/discriminative
  Answer:
  Hidden Markov model is **generative model** where the joint probability distribution of latent and observed variables are modeled whereas **Conditional Random Field** is a **discriminative model** where conditional probability is calculated.

- Objective function optimized
  Answer:
  HMM Optimizes joint likelihood where as, CRF optimizes Conditional probability function.

- Require a normalization constant
  Answer:
  CRF requires Normalization every step, however HMM doesn't since we are already calculating joint probability distribution.

3 Hidden Markov Model [50 pts]

This problem will let you get familiar with HMM algorithms by doing the calculations by hand.
[a-c] There are three coins $(1, 2, 3)$, to throw them randomly, and record the result. $S = 1, 2, 3$; $V = H, T$ (Head or Tail); A, B, π is given as

A:

	1	2	3
1	0.9	0.05	0.05
2	0.45	0.1	0.45
3	0.45	0.45	0.1

π:

π	1/3	1/3	1/3

B:

	1	2	3
H	0.5	0.75	0.25
T	0.5	0.25	0.75

(a) Given the model above, what's the probability of observation $O = H, T, H$. [10 pts]

Solution:
Initialization:

$$\alpha_1^k = P(x_1, y_1^k = 1) = P(x_1|y_1^k = 1)\phi_k$$

$$\alpha_1^1 = P(x_1 = H|y_1^1 = 1)\phi_1 = 0.5 * 1/3 = 1/6$$

$$\alpha_1^2 = P(x_1 = H|y_1^2 = 1)\phi_2 = 0.75 * 1/3 = 1/4$$

$$\alpha_1^3 = P(x_1 = H|y_1^3 = 1)\phi_3 = 0.25 * 1/3 = 1/12$$

Iteration:

$$\alpha_t^k = P(x_t|y_t^k = 1)\sum_i \alpha_{t-1}^i a_{i,k}$$

$$\alpha_2^1 = P(x_2 = T|y_2^1 = 1)\sum_i \alpha_1^i a_{i,1}$$

$$\alpha_2^1 = 0.5[1/6 * 0.9 + 1/4 * 0.45 + 1/12 * 0.45] = 0.15$$

$$\alpha_2^2 = P(x_2 = T|y_2^2 = 1)\sum_i \alpha_1^i a_{i,2}$$

$$\alpha_2^2 = 0.25[1/6 * 0.05 + 1/4 * 0.1 + 1/12 * 0.45] = 0.0177083$$

$$\alpha_2^3 = P(x_2 = T|y_2^3 = 1)\sum_i \alpha_1^i a_{i,3}$$

$$\alpha_2^3 = 0.75[1/6 * 0.05 + 1/4 * 0.45 + 1/12 * 0.1] = 0.096875$$

$$\alpha_3^1 = P(x_3 = T|y_3^1 = 1) \sum_i \alpha_2^i a_{i,1}$$

$$\alpha_3^1 = 0.5[0.15 * 0.9 + .0177083 * 0.45 + .096875 * 0.45] = 0.09328$$

$$\alpha_3^2 = P(x_3 = T|y_3^2 = 1) \sum_i \alpha_2^i a_{i,2}$$

$$\alpha_3^2 = 0.75[0.15 * 0.05 + .0177083 * 0.1 + .096875 * 0.45] = 0.03965$$

$$\alpha_3^2 = P(x_2 = T|y_2^2 = 1) \sum_i \alpha_2^i a_{i,3}$$

$$\alpha_3^2 = 0.25[0.15 * 0.05 + .0177083 * 0.45 + .096875 * 0.1] = 0.00629$$

Termination:

$$P(X) = \sum_k \alpha_T^k = 0.09328 + 0.03965 + 0.00629 = 0.13921875$$

Therefore Probability of Observation H,T,H is 0.13921875

(b) Describe how to get the A, B, and π, when they are unknown. [10 pts]

Solution:
We use Baum Welch Algorithm to learn the parameters A,B and π from X. Since Y is hidden/unknown, we use EM algorithm to converge on Y, A,B,π, for each observation.
(Step 1:) Initialize A,B,π to random values. EM algorithm:

E step

(Step2:) We calculate the partial probabilities using Forward backward algorithm as seen in the lecture. Where, beta is calculated using backward algorithm,

$$\beta_T^k = 1$$

$$\beta_t^k = \sum_i a_{k,i} b_{t+1} \beta_{t+1}^i$$

Alpha is calculated using forward algorithm,

$$\alpha_1^k = P(x_1|y_1^k = 1)\pi_k$$

$$\alpha_t^k = P(x_t|y_t^k = 1) \sum_i \alpha_{t-1}^i a_{i,k}$$

We use above results to calculate following for each sample of x observed:

$$P(x_n) = \sum_k \alpha_1^k \beta_1^k$$

$$\gamma_{n,t}^i = p(y_{n,t}^i = 1|X_n) = \frac{\alpha_t^i \beta_t^i}{P(x_n)}$$

$$\xi_{n,t}^{i,j} = p(y_{n,t-1}^i = 1, y_{n,t}^j = 1|X_n)$$

$$= \frac{P(X_n|y_{n,t-1}^i = 1, y_{n,t}^j = 1)P(y_{t-1}^i, y_t^j)}{P(X_n)} = \frac{P(X_{n,1}x_{n,2}...x_{n,t-1}|y_{n,t-1}^i)P(X_{n,t}x_{n,t+1}...x_{n,T}|y_{n,t}^j)P(y_{t-1}^i, y_t^j)}{P(X_n)}$$

$$P(X_{n,1}x_{n,2}...x_{n,t-1}|y_{n,t-1}^i) = \frac{P(y_{n,t-1}^i, X_{n,1}x_{n,2}...x_{n,t-1})}{P(y_{n,t-1}^i = 1)} = \frac{\alpha_{t-1}^i}{P(y_{n,t-1}^i = 1)}$$

$$P(X_{n,t}x_{n,t+1}...x_{n,T}|y_{n,t}^j) = \beta_t^j P(X_{n,t}|y_{n,t}^j) = \beta_t^j b_{jt}$$

Substituting, we get

$$\xi_{n,t}^{i,j} = \frac{\alpha_{t-1}^i \beta_t^j b_{jt} a_{ij}}{\sum_k \alpha_1^k \beta_1^k}$$

M step

$$\pi_i = \frac{\sum_n \gamma_{n,t}^i}{N}$$

$$\alpha_{ij} = \frac{\sum_n \sum_{t=2}^{\tau} \xi_{n,t}^{i,j}}{\sum_n \sum_{t=1}^{\tau-1} \gamma_{n,t}^i},$$

$$b_{ik} = \frac{\sum_n \sum_{t=1}^{\tau} \gamma_{n,t}^i x_{n,t}^k}{\sum_n \sum_{t=1}^{\tau-1} \gamma_{n,t}^i},$$

Note that $\alpha_{i,j}$ is members of matrix A, $b_{i,k}$ members of matrix B and π_i are values of vector π.

(c) In class, we studied discrete HMMs with discrete hidden states and observations. The following problem considers a continuous density HMM, which has discrete hidden states but continuous observations. Let $S_t \in 1, 2, ..., n$ denote the hidden state of the HMM at time t, and let $X_t \in R$ denote the real-valued scalar observation of the HMM at time t. In a continuous density HMM, the emission probability must be parameterized since the random variable X_t is no longer discrete. It is defined as $P(X_t = x|S_t = i) = \mathcal{N}(\mu_i, \sigma_i^2)$. Given m sequences of observations (each of length T), derive the EM algorithm for HMM with Gaussian observation model. [14 pts]

Solution:
Reference, Bishop text book
We know that the function we try to optimize for optimizing the expectation is,

$$Q(\theta, \theta^{old}) = \sum_z p(S|X, \theta^{old}) ln(p(X, S|\theta))$$

$$(p(X,S|\theta) = \prod_{d=1}^{D} (\pi_{s_1^{(d)}} B_{s_1^{(d)}}(x_1^{(d)}) \prod_{t=2}^{T} A_{s_{t-1}^{(d)} s_t^{(d)}} B_{s_t^{(d)}}(x_t^{(d)}))$$

$$Q(\theta,\theta^{old}) = \sum_{s\in S}\sum_{d=1}^{D} log\pi_{s_{t-1}^{(d)}} P(S,X;\theta^{old}) + \sum_{s\in S}\sum_{d=1}^{D}\sum_{t=2}^{T} logA_{s_{t-1}^{(d)} s_t^{(d)}} P(S,X;\theta^{old}) + \sum_{s\in S}\sum_{d=1}^{D}\sum_{t=1}^{T} logB_{s_t^{(d)}}(x_t^{(d)}) P(S,X;\theta^{o}$$

Please note that here B is the only term which is Gaussian and continuous. Rest all follow the same old EM algorithm and since differentiation remains the same as before.
(After applying Lagrangian and differentiating)

$$\hat{L}(\theta,\theta^{old}) = Q(\theta,\theta^{old}) - \lambda_\pi(\sum_{i=1}^{M}\pi_i - 1) - \sum_{i=1}^{M}\lambda_{A_i}(\sum_{j=1}^{M} A_{ij} - 1) - \sum_{i=1}^{M}\lambda_{B_i}(\sum_{j=1}^{N} B_i(j) - 1)$$

Differentiating w.r.t π, we get

$$\pi_i = \frac{\sum_{d=1}^{D} P(s_1^d = i|X^{(d)};\theta^{old})}{D}$$

Differentiating w.r.t A_{ij}, we get

$$A_{ij} = \frac{\sum_{d=1}^{D}\sum_{t=2}^{T} P(s_t^d = j, s_{t-1}^d = i|X^d,\theta^{old})}{\sum_{d=1}^{D}\sum_{t=2}^{T} P(s_{t-1}^d = i|X^d,\theta^{old})}$$

The above calculations are same as that of regualr Baum-Welch algorithm EM optimization. However B is differnet since there are additional parameters we need to optimize, which are μ_i and σ_i^2.

$$\frac{\partial L}{\partial \mu} = 0$$

$$=> \frac{\partial \sum_{s\in S}\sum_{d=1}^{D}\sum_{t=1}^{T} logB_{s_t^{(d)}}(x_t^{(d)}) P(S_t^d = i|X^d;\theta^{old})}{\partial \mu_i} = 0$$

$$=> \frac{\partial \sum_{s\in S}\sum_{d=1}^{D}\sum_{t=1}^{T} logN(\mu_i,\sigma_i^2)(x_t^{(d)}) P(S_t^d = i|X^d;\theta^{old})}{\partial \mu_i} = 0$$

$$=> \frac{\partial \sum_{s\in S}\sum_{d=1}^{D}\sum_{t=1}^{T} log(1/\sqrt{2\pi\sigma_i^2} e^{-(x_t^{(d)}-\mu_i)^2/(2\sigma_i^2)}) P(S_t^d = i|X^d;\theta^{old})}{\partial \mu_i} = 0$$

Expanding and ignoring σ and constant only terms, we get

$$=> \sum_{d=1}^{D}\sum_{t=1}^{T} ((x_t^{(d)} - \mu_i)/(2\sigma_i^2)) P(S_t^d = i|X^d;\theta^{old}) = 0$$

$$=> \sum_{d=1}^{D}\sum_{t=1}^{T} ((x_t^{(d)} - \mu_i) P(S_t^d = i|X^d;\theta^{old}) = 0$$

$$\boxed{=> \mu_i = \frac{\sum_{d=1}^{D}\sum_{t=1}^{T} (x_t^{(d)} P(S_t^d = i|X^d;\theta^{old}))}{\sum_{d=1}^{D}\sum_{t=1}^{T} P(S_t^d = i|X^d;\theta^{old})}}$$

$$\frac{\partial L}{\partial \sigma_i} = 0$$

$$=> \frac{\partial \sum_{s \in S} \sum_{d=1}^{D} \sum_{t=1}^{T} log(1/\sqrt{2\pi\sigma_i^2} e^{-(x_t^{(d)} - \mu_i)^2/(2\sigma_i^2)}) P(S_t^d = i | X^d; \theta^{old})}{\partial \sigma_i} = 0$$

$$=> \frac{\partial \sum_{s \in S} \sum_{d=1}^{D} \sum_{t=1}^{T} (log(1/\sqrt{2\pi}) - log(\sigma_i) - ((x_t^{(d)} - \mu_i)^2/(2\sigma_i^2))) P(S_t^d = i | X^d; \theta^{old})}{\partial \sigma_i} = 0$$

$$=> \sum_{d=1}^{D} \sum_{t=1}^{T} (-1/(\sigma_i) + ((x_t^{(d)} - \mu_i)^2/(\sigma_i^3))) P(S_t^d = i | X^d; \theta^{old}) = 0$$

Multiplying by σ_i^3 on both sides, we get

$$=> \sum_{d=1}^{D} \sum_{t=1}^{T} (((x_t^{(d)} - \mu_i)^2) - \sigma_i^2) P(S_t^d = i | X^d; \theta^{old}) = 0$$

$$\boxed{=> \sigma_i^2 = \frac{\sum_{d=1}^{D} \sum_{t=1}^{T} ((x_t^{(d)} - \mu_i)^2) P(S_t^d = i | X^d; \theta^{old})}{\sum_{d=1}^{D} \sum_{t=1}^{T} P(S_t^d = i | X^d; \theta^{old})}}$$

Now use the above recursion for EM algorithm.

(d) For each of the following sentences, say whether it is true or false and provide a short explanation (one sentence or so). [16 pts]

- The weights of all incoming edges to a state of an HMM must sum to 1.
  false, the weights of all incoming edges to a state of an HMM need not sum to 1.
  Explanation: In part a of this problem, the incoming edges to 1 sums up to 1.8 which not 1. This is due to the fact that the probabilities need not be independent. However the sum of weights of the edges going out of a state need to be 1 always.

- An edge from state s to state t in an HMM denotes the conditional probability of going to state s given that we are currently at state t. **false**, an edge from state s to state t in an HMM doesn't denote the conditional probability of going to state s given that we are currently at state t.
  It is reverse, the edge denotes conditional probability of going to state t given that we are currently at state s.

- The "Markov" property of an HMM implies that we cannot use an HMM to model a process that depends on several time-steps in the past.
  True, Markov property states that current state dependent on a single time step in the past.

- The Baum-Welch algorithm is a type of an Expectation Maximization algorithm and as such it is guaranteed to converge to the (globally) optimal solution.
  False, Baum-Welch algorithm converges only to a local maximum and not guaranteed to converge to the global maximum

4 Programming [30 pts]

In this problem, you will implement algorithm to analyze the behavior of *SP500* index over a period of time. For each week, we measure the price movement relative to the previous week and denote it using a binary variable (+1 indicates up and 1 indicates down). The price movements from week 1 (the week of January 5) to week 39 (the week of September 28) are plotted below.

Consider a Hidden Markov Model in which x_t denotes the economic state (good or bad) of week t and y_t denotes the price movement (up or down) of the *SP500* index. We assume that $x_{(t+1)} = x_t$ with probability 0.8, and $P_{(Y_t|X_t)}(y_t = +1|x_t = \text{good}) = P_{(Y_t|X_t)}(y_t = -1|x_t = \text{bad}) = q$. In addition, assume that $P_{(X_1)}(x_1 = \text{bad}) = 0.8$. Load the `sp500.mat`, implement the algorithm, briefly describe how you implement this and report the following :

(a) Assuming $q = 0.7$, plot $P_{(X_t|Y)}(x_t = \textbf{good}|y)$ for $t = 1, 2, ..., 39$. What is the probability that the economy is in a good state in the week of week 39. [15 pts]

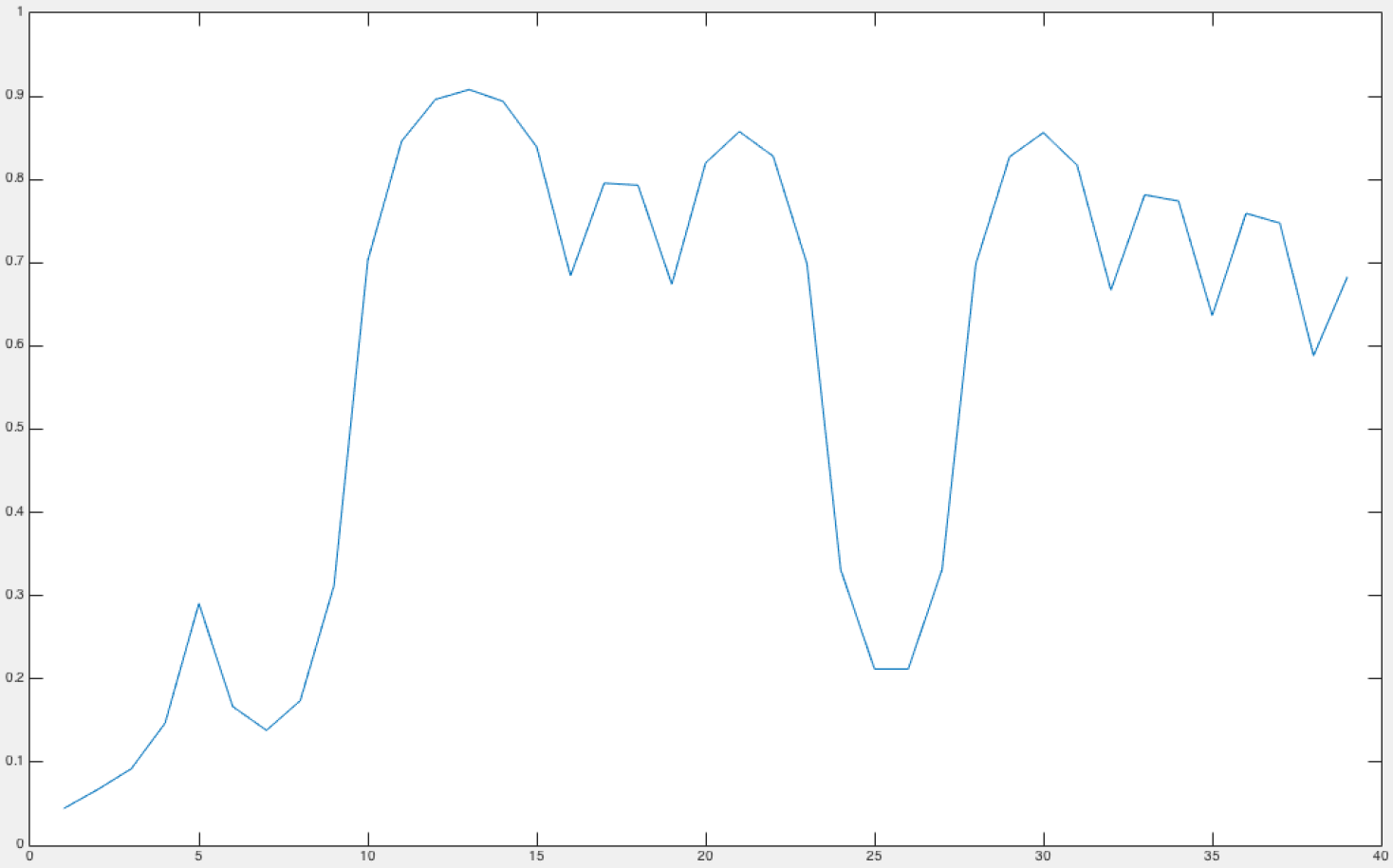
Solution
Let $p = P(x_{(t+1)} = x_t)$, we know that $P_{(Y_t|X_t)}(y_t = +1|x_t = \text{good}) = P_{(Y_t|X_t)}(y_t = -1|x_t = \text{bad}) = q$. Also $P_{(X_1)}(x_1 = \text{good}) = 0.8$
Using above equations, we can derive probability that $x_i = \text{good}$ given y_i using Bayes theorem.

$$P(x_t|y_t, y_{t-1}, ..., y1) = \frac{P(y_t|x_t, y_{t-1}, ..., y1)P(x_t|y_{t-1}, ..., y1)}{\sum P(y_t)}$$

We know that, given x_t, y_t is independent of $y_{t-1}, ..., y1$. Using this property,

$$P(x_t|y_t, y_{t-1}, ..., y1) = \frac{P(y_t|x_t)P(x_t|y_{t-1}, ..., y1)}{\sum P(y_t)}$$

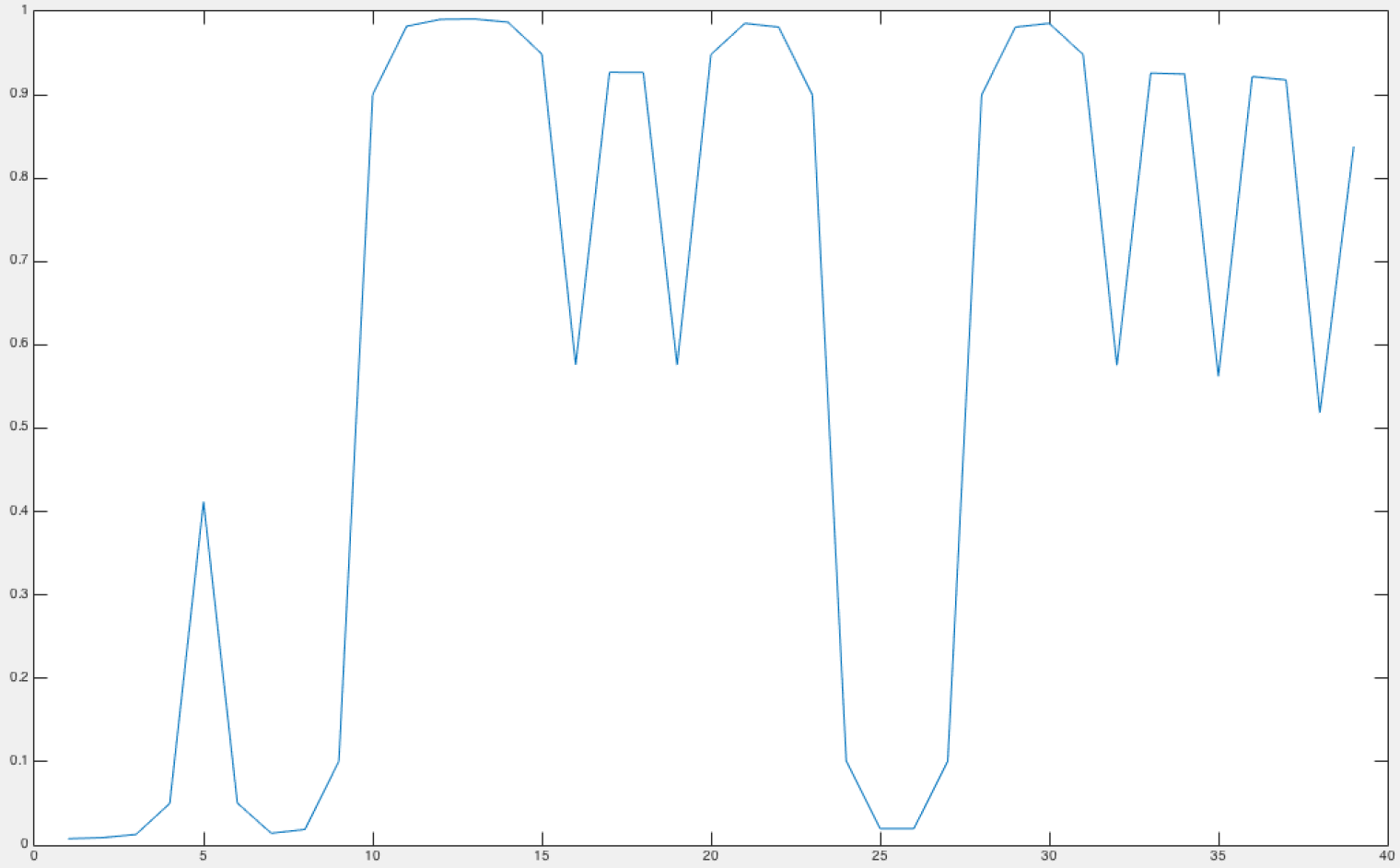
Also $P(x_t|y_{t-1}, ..., y1) = \sum P(x_{t-1}|y_{t-1}, ..., y1)P(x_t|P(x_{t-1})$
Using the above recursion, we are able to predict the probability of having good day on week 39 given all the input parameters.
Following is the graph for the above calculations obtained for the value,



probability that the economy is in a good state in the week of week 39 is **0.6830**

(b) Repeat (a) for $q = 0.9$, and compare the result to that of (a). Explain your comparison in one or two sentences. [15 pts]

Solution:
The graph for the probability prediction for 39 weeks is following



probability that the economy is in a good state in the week of week 39 is **0.8379**.

Comparison:
The rate of growth of probability towards 1 is some how proportional to q value. Hence the probability is 0.7 for first case while it is near 0.9 for the second one. This is due to the q value which clearly influences the numerator in probability calculation.
Also from the graph, probability fluctuation is more violent(even at bottom) in case of 0.9 again due to same reason compared to that of first graph (q=0.7).