CSE/ISYE 6740 Homework 4

Le Song

Deadline: 11/28 Thursday, 11:55 am

- **Please read the following submission rules carefully. We will strictly follow these rules starting from this assignment.**
- No unapproved extension of deadline is allowed. Late submission window will open immediately after the deadline for 12 hours, but with a -40 points penalty. No submission via email will be accepted.
- We recommend writing your report with Latex. Other text editors such as MS Word are also accepted. Hand-written report must be clear. Unreadable handwriting is subject to zero credit. Report must be **a single pdf** file named with the format **HW*number_your_GT_account*.pdf**. For example, "HW4_yyang123.pdf". Any additional document is ignored.
- For programming assignments, please strictly follow the provided function APIs, as we will be grading your code by running the homework2.m/homework2.py file with required arguments. You will be responsible for any runtime error due to the change of the API or code file name. You should implement your programs in the given .m/.py file. Any code pasted to the report or other files will not be graded.
- Matlab submissions will be graded in Matlab Online R2019b environment[1][2]. Python submission will be graded in Python 3.6 environment on a regular laptop. Please make sure your program is runnable in those environments. Any program that cannot run is subject to zero credit. Situations include but not restricted to exceeding 5 minutes of running time, exceeding memory limit and invalid text character.
- Put your report and code files into a single flat folder and **submit it as a single .zip file**.
- Explicitly mention your collaborators if any.
- Recommended reading: PRML Section 13.2

1 Kernels [20 points]

(a) 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. [8 pts]

Suppose K_1 and K_2 are valid kernels (symmetric and positive definite) defined on $R^m \times R^m$.

我不觉得symmetry是必须的

semi

1. $K(u,v) = \alpha K_1(u,v) + \beta K_2(u,v), \alpha, \beta \in R$.
2. $K(u,v) = K_1(f(u), f(v))$ where $f : R^m \to R^m$. coefficients.

[1]https://matlab.mathworks.com/

[2]You can also use the environment for developing by registering online with your university email.

3.

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

4. Suppose K' is a valid kernel.

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

(b) Write down kernelized version of Fisher's Linear Discriminant Analysis using kernel trick. Please provide full steps and all details of the method. [*Hint: Use kernel to replace inner products.*] [12 pts]

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]

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

(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
- Objective function optimized
- Require a normalization constant

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]

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

(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]

(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.
- 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.
- 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.
- 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.

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]

(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]