HW2, CSE 569 Fall 2019

Due 11:59pm Oct 20, 2019

Maximum score: 70 + Extra Credit 30

Q.1) (**20 pt**) The Poisson distribution is a discrete probability distribution that expresses the probability of a certain number of events (x) that occur in a given interval of time or space. The events are understood to be independent and the rate of occurrence is a constant given by λ. For e.g., the number of meteorites hitting the moon surface on a given moon day can be modeled by a Poisson distribution. The Poisson distribution is given by,

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}. \tag{1}$$

The time interval is one moon day. λ is the expected number of occurrences in the given interval and the events are independent - we assume a meteorite event does not depend on other meteorite events.

a) (**5 pt**) Seismometers on the moon estimate the number of meteorites hitting the moon surface on any given moon day. The following readings were observed over 5 days, $\{45, 36, 15, 25, 40\}$. If the number of meteorites hitting the moon on a moon day follows a Poisson distribution, (Eq. 1), calculate the value of λ using maximum likelihood. Outline the steps to your solution.

b) (**5 pt**) The conjugate prior for the Poisson distribution is the Gamma distribution, given by,

$$p(\lambda;\alpha,\beta) = \frac{\beta^\alpha \lambda^{\alpha-1}\exp(-\beta\lambda)}{\Gamma(\alpha)}, \qquad \lambda > 0 \quad \alpha,\beta > 0, \tag{2}$$

where, $\alpha, \beta > 0$ are parameters of the distribution and $\Gamma(.)$ is the Gamma function defined as $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$. The Gamma function $\Gamma(.)$, is different from the Gamma distribution. Show that,

$$\int_0^\infty \lambda^{\alpha-1}\exp(-\beta\lambda)d\lambda = \frac{\Gamma(\alpha)}{\beta^\alpha} \tag{3}$$

(Hint: $p(\lambda)$ is a distribution and will integrate to 1 over the space of λ)

c) (**10 pt**) Given a set of meteorite observations $D = \{x_1, x_2, \ldots, x_n\}$ on n moon days, estimate the posterior distribution $p(\lambda|D)$ using Bayes estimation,

$$p(\lambda|D) = \frac{p(D|\lambda)p(\lambda)}{p(D)},$$

where, $p(D|\lambda)$ is the maximum likelihood with $p(x;\lambda)$ being the Poisson distribution given by Eq. 1. $p(\lambda)$ is the prior distribution given by Eq. 2 and $p(D) = \int_0^\infty p(D|\lambda)p(\lambda)d\lambda$. Show that the posterior distribution $p(\lambda|D)$ is also a Gamma distribution $p(\lambda;\alpha^*,\beta^*)$. Estimate the new parameters α^* and β^* of the posterior

Gamma distribution. (Hint: the result from part (b) in Eq. 3 will be useful to simplify $p(D)$ in the posterior distribution.)

Q.2) (**20 pt**) Download the MNIST dataset from the website `http://yann.lecun.com/exdb/mnist/`. Using the functions `load_mnist.py` or `getMnistData.m`, sample 200 images each from digit 5 and digit 8 from the train set in MNIST to create a training set of 400 images. Sample, 50 images each from digits 5 and 8 from the test set in MINST to create a testing set of 100 images. We will work with these set of sampled images for the rest of the question. Apply PCA on the train data to reduce the dimension from 784 to 10. Apply the same transformation to the test data to reduce its dimension to 10.

1) Plot the covariance matrix for the PCA transformed train data. You can use `imagesc` in Matlab or `matplotlib.pyplot.matshow` in Python or similar plotting functions in other languages.
2) Reconstruct the original images from the PCA transformed data ($10 \to 784$ dimension) and display 5 images each from both the classes in two rows. Display the original images (before PCA) in order to compare them with the PCA reconstructed images.
3) Implement Fishers Linear Discriminant to project the PCA train data to 1 dimension and estimate a threshold to separate the two categories. Estimate the training accuracy and the test accuracy with the Fishers Linear Discriminant.

Outline the code as part of your answer in the report. DO NOT place an image of your code - typeset the code.

Q.3) (**30 pt**) Consider a Hidden Markov Model (HMM) with hidden states $\{1, 2, 3, 4\}$ where 1 is a stop hidden state (state with no external transition) and observed states $\{S, A, B, C, D\}$, where S is a Stop symbol. Consider the following 2 HMMs $\lambda_i = \{\pi_i, \mathbf{A}_i, \mathbf{B}_i\}$, for $1 \le i \le 2$ where,

$$\pi_1 = \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \atop \begin{pmatrix} 0.0 & 1.0 & 0.0 & 0.0 \end{pmatrix} \quad \mathbf{A}_1 = \begin{array}{c|c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.4 & 0.3 & 0.3 \\ 0.3 & 0.2 & 0.2 & 0.3 \end{pmatrix} \end{array} \quad \mathbf{B}_1 = \begin{array}{c|c} & \begin{array}{ccccc} S & A & B & C & D \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.2 & 0.3 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \end{pmatrix} \end{array}$$

$$\pi_2 = \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \atop \begin{pmatrix} 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix} \quad \mathbf{A}_2 = \begin{array}{c|c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.3 & 0.5 & 0.1 \\ 0.1 & 0.4 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0.2 & 0.3 \end{pmatrix} \end{array} \quad \mathbf{B}_2 = \begin{array}{c|c} & \begin{array}{ccccc} S & A & B & C & D \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.0 & 0.5 \end{pmatrix} \end{array}$$

Answer the following questions. Outline the code used to solve these problems. DO NOT put an image of the code - typeset the code.

a) (**10 pt**) Write procedure to generate 10 sequences of observations from HMM with λ_1 and list the observations with comma separated values. For e.g., two sequences from a HMM with observed states $\{S, A, B, C, D\}$ would be,

1) A, A, D, C, A, C, B, S
2) B, D, S

b) (**10 pt**) Implement the Forward algorithm to classify the following sequences by estimating $\log p(O|\lambda_i)$

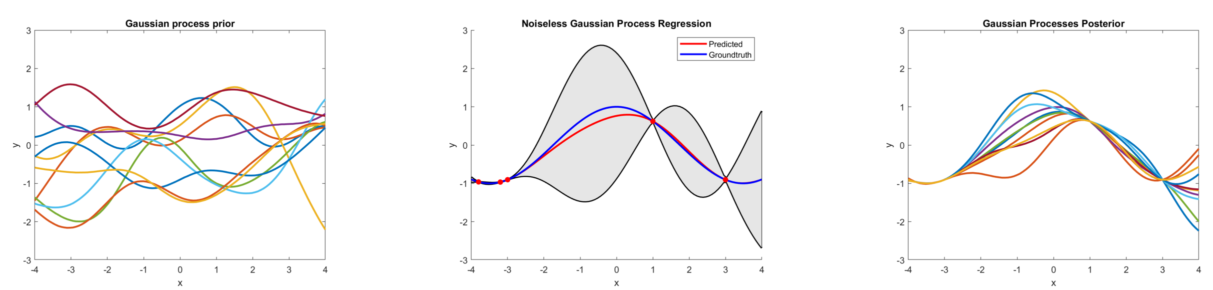
1) A, D, C, B, D, C, C, S
2) B, D, S
3) $B, C, C, B, D, D, C, A, C, S$
4) A, C, D, S
5) A, D, A, C, S
6) D, B, B, S
7) A, B, S
8) $D, D, B, D, D, B, A, C, C, D, A, B, B, C, D, B, B, B, S$
9) D, B, D, S
10) A, A, A, A, D, C, B, S

c) (**10 pt**) Implement the Viterbi algorithm to decode the hidden states for the above observations using HMM λ_2.

Q.4) (**30 pt - Extra Credit**) Implement non-linear regression using zero mean noiseless Gaussian Processes to estimate the function values $f(x) = 0.5 * \sin(x)$ for $x \in [-4, 4]$. Assume you are given training data, $D = \{(x_i, y_i)_{i=1}^5\} = \{(-3.8, -0.9463), (-3.2, -0.9996), (-3, -0.9975), (1, 0.4794), (3, 0.9975)\}$. Use the kernel function,

$$k(x, x') = \sigma_f^2 \exp\Big[\frac{1}{2l^2}(x - x')^2\Big],$$

with $\sigma_f^2 = l^2 = 1$.



(a) prior (b) Regression (c) Posterior

Fig. 1: Gaussian Processes Nonlinear Regression sample plots (a) Prior, (b) Nonlinear Regression, (c) Posterior

1) Plot 10 prior functions in the interval $[-4, 4]$ - see an example plot in Fig. 1(a).
2) Plot the mean estimate for the non-linear regression and the error curves above and below indicating confidence in the estimate - see an example plot in Fig. 1(b).
3) Plot 10 posterior functions sampling the conditional distribution $GP(f|D)$ - see an example plot in Fig. 1(c).

Outline the code. DO NOT put an image of the code - typeset the code.