HW2, CSE 569 Fall 2018

Due 11:59pm Oct 9, 2018

Maximum score: 6

You may answer any one of the following 3 questions (Q.I or Q.II or Q.III). Submit your report as a pdf. Append your code in the pdf if you are attempting either Q.II or Q.III. You get 1 point extra credit for every extra question you complete successfully. So, you can score a maximum of 8 points in this assignment. Write your own code and do not use library functions for HMM. You can use library functions (like `svd()`) for estimating eigenvectors and eigenvalues.

Q.I) (**6 pt**) MLE and Bayesian Learning.

(a) (2 pt) Consider a univariate random variable y belonging to a normal distribution $p(y|\boldsymbol{x}, \theta, \sigma^2) = \mathcal{N}(\boldsymbol{x}^\top \theta, \sigma^2)$. Here, $\boldsymbol{x} \in \mathbb{R}^d$ is a random variable which is given, θ and σ^2 are the parameters of the distribution. Apply Maximum-likelihood to estimate θ and σ^2. Note - Every y_i has a unique $\boldsymbol{x}_i$, $1 \leq i \leq n$.

(b) (4 pt) Let θ be the probability of tossing a head in a coin experiment. The prior probability distribution for θ is a given by the Beta distribution for $0 \leq \theta \leq 1$,

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

where, $\Gamma(z)$ is the Gamma function defined as $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ and α and β are hyper parameters of the Beta distribution.

(i) What is $\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$ (Hint: $p(\theta)$ is a distribution over θ)

(ii) Show that the mode of the distribution (value of θ where the distribution peaks) is given by $\frac{\alpha-1}{\alpha+\beta-2}$

(iii) Since the coin toss is a Bernoulli random variable, the likelihood of n events is given by $p(\mathcal{D}|\theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{n-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$. Show that the posterior distribution over θ is a Beta distribution and determine its hyper parameters.

(iv) What is the mode of the posterior distribution. Analyze the posterior when the hyper parameters of the prior distribution are $\alpha = \beta = 1$.

Q.II) (**6 pt**) PCA and Fishers Linear Discriminant. For this question, you will work with the MNIST [1] dataset. Download the MNIST dataset (help code is provided in python and Matlab). Select only 1000 training digit images (100 per each digit). Select 100 test digit images (10 per digit). Submit your report and code in the form of a pdf.

[1] http://yann.lecun.com/exdb/mnist/

(a) (2 pt) Perform PCA and reduce the dimensions of training images from $28 \times 28 = 784$ to 100. Output the percentage of variance retained in the data after dimensionality reduction. Plot the covariance matrix for the training data after dimensionality reduction and analyze. When plotting covariance matrices, use `imagesc` in Matlab or `matplotlib.pyplot.matshow` in case of Python or similar functions for other libraries. Plot one example of each of the digits after dimensionality reduction and analyze.

(b) (2 pt) Perform PCA whitening on the reduced dimensions. When implementing PCA whitening, we set $e_i := e_i/\sqrt{\lambda_i}$. If $\lambda_i \to 0$, this can lead to numerical instability. In such cases, implement regularization where you set $e_i := e_i/\sqrt{\lambda_i + \epsilon}$ and let $\epsilon = 1e-3$. Plot the covariance matrix after whitening and analyze. Plot one example of each of the digits after PCA whitening and analyze. Implement ZCA whitening and plot the covariance matrix. Plot one example of each of the digits after ZCA whitening and analyze.

(c) (2 pt) Implement Fishers Linear discriminant for digits 0 and 1 in the training dataset. Use reduced PCA dimensions (100). Report the accuracy on the test dataset having only digits 0 and 1.

Q.III) (**6 pt**) Hidden Markov Models. For this question submit the report and code in the form of a pdf. Consider the use of hidden Markov models for classifying sequences of four visible states, `A-D`. Train two hidden markov models, each consisting of three hidden states (plus a null initial state and a null final state), fully connected, with the following data. Assume that each sequence starts with a null symbol and ends with a null symbol (not listed).

Sample	ω_1	ω_2
1	AABBCCDD	DDCCBBAA
2	ABBCBBDD	DDABCBA
3	ACBCBCD	CDCDCBABA
4	AD	DDBBA
5	ACBCBABCDD	DADACBBAA
6	BABAADDD	CDDCCBA
7	BABCDCC	BDDBCAAAA
8	ABDBBCCDD	BBABBDDDCD
9	ABAAACDCCD	DDADDBCAA
10	ABD	DDCAAA

(a) (2 pt) Print out the full transition matrices for each of the models.

(b) (2 pt) Assume equal prior probabilities for the two models and classify each of the following sequences: `ABBBCDDD`, `DADBCBAA`, `CDCBABA`, and `ADBBBCD`.

(c) (2 pt) As above, classify the test pattern `BADBDCBA`. Find the prior probabilities for your two trained models that would lead to equal posteriors for your two categories when applied to this pattern.