CS 539 Machine Learning Homework 5

Conceptual and Theoretical Questions (4 questions, 40 pts)

1. Evaluate the Kullback-Leibler divergence (equation 1.113 in Bishop) between two Gaussians $p(x) = N(x|\mu, \Sigma)$ and $q(x) = N(x|m, L)$. (This is question 2.13, from Bishop's textbook). **(10 pts)**
2. Suppose that $p(x)$ is some fixed distribution and that we wish to approximate it using a Gaussian distribution $q(x) = N(x|\mu, \Sigma)$. By writing down the form of the KL divergence $KL(p, q)$ for a Gaussian q(x) and then differentiating, show that minimization of KL(p,q) with respect to μ and Σ leads to the result that μ is given by the expectation of x under $p(x)$ and that Σ is given by the covariance. (This is question 10.4, from Bishop's textbook) **(10 pts)**
3. Consider a regression problem involving multiple target variables in which it is assumed that the distribution of the targets, conditioned on the input vector x, is a Gaussian of the form
$$p(t|x, w) = N(t|y(x, w), \Sigma) \qquad (5.192)$$
where $y(x, w)$ is the output of a neural network with input vector x and weight vector w, and Σ is the covariance of the assumed Gaussian noise on the targets.
Given a set of independent observations of x and t, write down the error function that must be minimized in order to find the maximum likelihood solution for w, if we assume that Σ is fixed and known. Now assume that Σ is also to be determined from the data, and write down an expression for the maximum likelihood solution for Σ. Note that the optimizations of w and Σ are now coupled, in contrast to the case of independent target variables discussed in Section 5.2.
(This is question 5.3, from Bishop's textbook) (**10 pts**)
4. Let $z = e^{x^2 y}$, where $x(u;\ v) = \sqrt{uv}$ and $y(u;\ v) = 1/v$. **(10 pts)**
    a. Derive $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$.
    b. Let's assume the target value for the output (z) is t. We want to minimize the $e = \frac{1}{2}(t - z)^2$; write down the update rule for changing u and v that minimizes e.

Application Questions (3 questions, 60 pts)

KL Distance (20 points) In question 1 in the theoretical part, we derived the KL distance between two multi-variate normal. Here, we want to study KL distance for a more complex distribution. Let's assume

$$p(x) = \frac{1}{3}\, N(x; -1,2) + \frac{2}{3}\, N(x; 1,1) \qquad\qquad N(x; mean, variance)$$

and we want to approximate it by

$$q(x) = N(x; m_a, \sigma_m^2)$$

by minimizing the KL distance between $p(x)$ and $q(x)$ – $KL(p,q)$.

a. Show KL distance for different values of m_a and σ_m^2. To do this, you can pick different values for m_a and σ_m^2, and use sampling techniques to calculate KL distance.
b. In question 2, we provided the answer for m_a and σ_m^2. Find the mean and variance for x given $p(x)$, and compare that we the result in part a.

TSNE (20 points)
We discussed TSNE in class, and we want to explore its application in a dataset. The UCI HAR Dataset has two folders: one training and one test folder. For the test folder, the X_test.txt includes the attributes (features), y_test.txt carries the activities' labels, and subject_test.txt carries participants' indexes. You can check README.txt to get more information about the dataset. There are 30 participants and 6 activities in the data. For this problem, use TSNE toolset to visualize possible manifold present in your data.

a) Run TSNE on the training set and discuss the visualization results. Here, you need to check how the visualization represents participants or activities (**Figure 1** shows an example TSNE result).

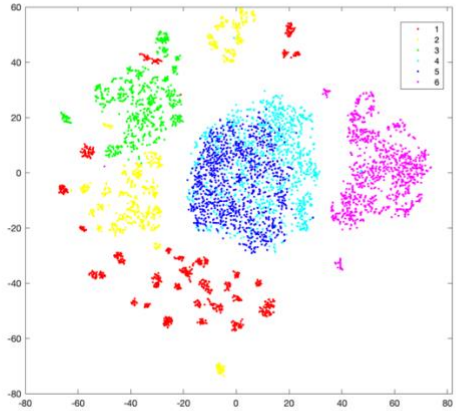

Figure 1. Example TSNE visualization. Different colors represent different activities. I used Matlab tsne function to create this figure.

b) Run TSNE on the test set and discuss the visualization results.
c) Discuss possible similarities (dissimilarities) you will see in parts *a* and *b*.

Note: you can use different distance metrics to reach better separations. Note that in practice, you might not have the labels.
Link to the dataset:
https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Neural Networks (20 points)

We want to build NN to classify different activities and participants given their activities. The UCI HAR Dataset has two folders: one training and one test folder. For the test folder, the X_test.txt includes the attributes (features), y_test.txt carries the activities' labels, and subject_test.txt carries participants' indexes. You can check README.txt to get more information about the dataset. There are 30 participants and 6 activities in the data.

a) Build a NN model to classify activities given feature. Discuss your NN topology, and your test and training performance. Note the dataset provides test and training folders.

b) Build a NN model to recognize participants. Discuss your NN topology, and your test and training performance. *Note that we have two separate groups of participants in the test and training dataset; as a result, for this question, we first concatenate the test and training datasets and use 10-fold cross-validation to assess the classification performance.* Note the dataset provides test and training folders.

Note: I used the fitcnet function for this problem. You can explore different topologies and provide comparison results as well.

Link: https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones