Mid Semester Examination

Session: 2024-2025

Semester: Monsoon

Time: 2 Hours

NECD509: Machine Learning

Maximum Marks: 64

1. Let us consider a housing price prediction scenario where we want to predict the price of the house on the basis of single input feature, i.e, size of the house. Further, we decide to fit the training data of size N of housing price against the size of each house with the polynomial function of the form given as $\sum_{j=0}^{M} \theta_j x^j$, where M is the polynomial order, x represents the size of the house. and θ_j is the weight associated with the j^{th} order polynomial. Furthermore, we are unable to gather more training data and thus, we have the situation where $N = M = 10$. With this, we move on to run the gradient descent algorithm so as to minimize the least square cost function. Based on this, answer the following

   (a) What trend can we expect for the training error ? Can we expect our trained algorithm would make good predictions on unseen testing data? (4 marks)

   (b) Mention the necessary modifications in the cost function if the trained algorithm works very badly on the testing data set. (4 marks)

   (c) Based on the modified cost function, derive the update rule for θ's considering both the batch gradient descent as well as stochastic gradient descent algorithm. (10 marks)

   (d) Without resorting to an iterative algorithm, derive the optimum vector $\bar{\theta}_{\text{optimum}}$ that minimizes the modified cost function. (4 marks)

2. Suppose that you have the following training data set : $\left\{ \left(\bar{x}^{(1)}, z^{(1)}\right), \left(\bar{x}^{(2)}, z^{(2)}\right), \ldots, \left(\bar{x}^{(n)}, z^{(n)}\right) \right\}$. Further, we have $\bar{x}^{(i)} \in \mathbb{R}^d$ and $z^{(i)} \in \mathbb{R}$ for $i \in \{1, 2, \ldots, n\}$. We are trying to fit the training data with the linear function of the form $\sum_{j=0}^{d} \theta_j x_j = \left(\bar{\theta}\right)^T \bar{x}$. Furthermore, we have the following relation

$$z^{(i)} = \left(\bar{\theta}\right)^T \bar{x}^{(i)} + w^{(i)}, \quad i \in \{1, 2, \ldots, n\}$$

where each $w^{(i)}$ is independent and identically distributed (IID) Gaussian with zero mean and unit variance. The notation $(\cdot)^T$ denotes the transpose operation. Find the maximum likelihood estimate of $\bar{\theta}$. (10 marks)

3. Let us consider the following training data set : $\{(\bar{x}^{(1)}, y^{(1)}), (\bar{x}^{(2)}, y^{(2)}), \ldots, (\bar{x}^{(m)}, y^{(m)})\}$, i.e., we have m independent training examples. Further, we have $\bar{x}^{(i)} \in \mathbb{R}^n$ and $y^{(i)} \in \{0, 1\}$ for $i \in \{1, 2, \ldots, m\}$. Based on this, consider the following tasks

(a) For the aforementioned training data, suggest the appropriate hypothesis function. (2 marks)

(b) Derive the cost function expression for the problem and thereafter, compute the stochastic gradient descent rule for updation of the weight parameters, i.e. θ's. (10 marks)

4. (a) Let us consider a Poisson distributed discrete random variable Y with probability mass function (PMF) given by

$$f(y; \lambda) = \mathbb{P}(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, \lambda > 0, \ \mathbb{E}[Y] = \lambda, \text{ and } y = 0, 1, 2, \ldots,$$

Here $\mathbb{E}[\cdot]$ denotes the Expectation operator. Show that the random variable Y falls under the category of exponential distribution. (5 marks)

(b) Considering the regression using the generalised linear models (GLMs) with Poisson distributed response variable, i.e., the output labels are Poisson distributed. compute the canonical response function. (5 marks)

(c) Considering m independent training examples $\{(\bar{x}^{(1)}, y^{(1)}), (\bar{x}^{(2)}, y^{(2)}), \ldots, (\bar{x}^{(m)}, y^{(m)})\}$ where $f(y^{(i)}|\bar{x}^{(i)}; \bar{\theta})$ is Poisson distributed and using the canonical response function derived in (b), derive the stochastic gradient ascent update rule for $\bar{\theta}$. (10 marks)