Question Paper

Exam Date & Time: 13-Jul-2022 (09:00 AM - 12:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SIXTH SEMESTER B.TECH MAKEUP EXAMINATIONS, JULY 2022

MACHINE LEARNING [ICT 4032]

Marks: 50 **Duration: 180 mins.**

A

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) A) The marginal distributions of Gaussians are themselves Gaussians, and as per the definition of the multivariate Gaussian distribution, it is known that $x_1|x_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$, where (5)

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma^{-1}(x_2 - \mu_2)$$
$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma^{-1}\Sigma_{21}$$

In a factor analysis model, assume a joint distribution on (x, z) as follows

$$z \sim \mathcal{N}(0, I)$$
$$x|z \sim \mathcal{N}(\mu + \Lambda z, \Psi)$$

where $\mu \in \mathbb{R}^n$, $\Lambda \in \mathbb{R}^{n \times k}$, and the diagonal matrix $\Psi \in \mathbb{R}^{n \times n}$, $(k < n)$. Derive the expression for the log likelihood of the parameters $l(\mu, \Lambda, \Psi)$.

B) For ordinary least squares problem, we fit $\theta^T x$, and the batch gradient update is given by (3)

$$\theta := \theta + \alpha \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)}).$$

Consider $\phi : \mathbb{R}^n \to \mathbb{R}^p$ be a feature map that maps attribute $x \in \mathbb{R}^n$ to factor $\phi(x) \in \mathbb{R}^p$. Apply the kernel trick to the batch gradient and derive the result for the parameter update.

C) Define the following: (2)

i. Empirical error

ii. Empirical risk minimization.

2) Describe various types of ambiguities in context of independent component analysis. (5)

A)

B) Describe the following event models for text classification: (3)

i. Multi-variate Bernoulli event model

ii. Multinomial event model.

C) Assume that the input feature $x_j,\ j = 1, \ldots, n$ are discrete binary-valued variables such that $x_j \in \{0, 1\}$ and $x = [x_1 x_2, \ldots, x_n]$. For each training example $x^{(i)}$, assume that the output target variable $y^{(i)} \in \{0, 1\}$. Now, consider the Naive Bayes model, given the above context. This model can be parametrized by $\phi_{j|y=0} = p(x_j = 1|y = 0)$, $\phi_{j|y=1} = p(x_j = 1|y = 1)$, and $\phi = p(y = 1)$. Write the expression for $p(y = 1|x)$ in terms (2)

of $\varphi_{j|y=0}, \varphi_{j|y=1}$, and φ_y.

3) A) Suppose you are given a dataset $\{(x^{(i)}, y^{(i)}; i = 1, \ldots, m)\}$ consisting of m independent examples, where $x^{(i)} \in \mathbb{R}^n$ are n-dimensional vectors, and $y^{(i)} \in \{0, 1\}$. You will model the joint distribution of (x, y) according to: (5)

$$p(y) = \phi^y (1-\phi)^{1-y}$$
$$p(x|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\Big).$$
$$p(x|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\Big)$$

Suppose you have already fit ϕ, μ_0, μ_1, and Σ, and now want to make a prediction at some new query point x. Show that the posterior distribution of the label at x takes the form of a logistic function, and can be written as

$$p(y = 1|x; \phi, \Sigma, \mu_0, \mu_1) = \frac{1}{1 + \exp(-\theta^T x)},$$

where θ is some appropriate function of $\phi, \Sigma, \mu_0, \mu_1$.

B) Describe the following: (3)

i. Cross validation

ii. K-fold cross validation

iii. Leave-one-out cross validation.

C) Write the algorithms for value iteration and policy iteration. (2)

4) A) Assume that the target variable and the inputs are related via $y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$, where $\epsilon^{(i)}$ is an error term that captures either unmodeled effects or random noise. Further, assume that $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$, and the density of $\epsilon^{(i)}$ is given by (5)

$$p(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\epsilon^{(i)})^2}{2\sigma^2}}.$$

Using these probabilitic assumption on the data, show that the least-square regression corresponds to finding the maximum likelihood estimate of θ.

B) Consider a modified algorithm, called the Support Vector Regression algorithm, which can be used for regression with continuous valued labels $y \in \mathbb{R}$. Suppose we are given a training set $\{(x^{(i)}, y^{(i)}); i = 1, \ldots, m\}$, where $x^{(i)} \in \mathbb{R}^{n+1}$ and $y \in \mathbb{R}$. We would like to find a hypothesis of the form $h_{w,b}(x) = w^T x + b$ with a small value of w. Our optimization problem is (3)

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)} - w^T x^{(i)} - b \leq \epsilon, \ i = 1, \ldots, m \\ & w^T x^{(i)} + b - y^{(i)} \leq \epsilon, \ i = 1, \ldots, m \end{aligned}$$

where $\epsilon > 0$ is a given, fixed value.

i) Write the Lagrangian for the given optimization problem. Use two sets of Lagrange multipliers α_i and β_i, corresponding to the two inequality constraints, so that the Lagrangian would be written as $\mathcal{L}(w, b, \alpha, \beta)$.

ii) Derive the dual optimization problem

ii) Derive the dual optimization problem.

C) Briefly explain the filtering method for feature selection. (2)

5) Explain the concept of functional and geometric margin in reference to Support Vector Machine (SVM). Pose an optimization problem in terms of geometric margin such that its solution gives the optimal margin classifier. (5)

A)

B) The regularized least squares has cost function (3)

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})^2 + \frac{\lambda}{2}||\theta||^2.$$

Use the vector notation to find a closed-form expression for the value of θ which minimizes the given cost function.

C) Write the pre-processing steps before applying PCA. (2)

-----End-----