Question Paper

Exam Date & Time: 07-Jul-2023 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SIXTH SEMESTER B.TECH MAKEUP EXAMINATIONS, JULY 2023

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) (5)

A)

This problem is about maximum likelihood parameter estimation using the naive Bayes assumption. Here, the input features $x_j, j = 1, \ldots, n$ to our model are discrete, binary-valued variables, so $x_j \in \{0, 1\}$. We call $x = [x_1 \; x_2 \; \cdots \; x_n]^T$ to be the input vector. For each training examples, our output targets are a single binary-value $y \in \{0, 1\}$. Our model is parametrized by $\phi_{j|y=0} = p(x_j = 1|y = 0)$, $\phi_{j|y=1} = p(x_j = 1|y = 1)$, and $\phi_y = p(y = 1)$. We model the joint distribution of (x, y) according to

$$p(y) = (\phi_y)^y (1 - \phi_y)^{1-y}$$

$$p(x|y = 0) = \prod_{j=1}^{n} p(x_j|y = 0)$$

$$= \prod_{j=1}^{n} (\phi_{j|y=0})^{x_j} (1 - \phi_{j|y=0})^{1-x_j}$$

$$
\begin{aligned}
p(x|y=1) &= \prod_{j=1}^{n} p(x_j|y=1) \\
&= \prod_{j=1}^{n} (\phi_{j|y=1})^{x_j}(1-\phi_{j|y=1})^{1-x_j}
\end{aligned}
$$

i) Find the joint likelihood function $\ell(\varphi) = \log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \varphi)$ in terms of the model parameters given above. Here, φ represents the entire set of parameters $\{\phi_y, \phi_{j|y=0}, \phi_{j|y=1}, j = 1, \dots, n\}$.

ii) Using the results of (i), show that $\phi_{j|y=0} = \frac{\sum_{i=1}^{m} 1\{x_j^{(i)}=1 \wedge y^{(i)}=0\}}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}}$.

B) For logistic regression, derive the relation for stochastic gradient ascent rule. (3)

C) Consider the geometric distribution, which is parametrized by ϕ given by (2)

$$p(y;\phi) = (1-\phi)^{y-1}\phi.$$

Show that the geometric distribution is an exponential family distribution. Explicitly specify $b(y), \eta, T(y)$, and $a(\eta)$. Also write ϕ in terms of η.

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

A)

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 $\{(x^{(i)}, y^{(i)}); i = 1, \dots, m\}$, 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)

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

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.

C) Consider a Markov model with given set of states $S = \{s_1, s_2, \dots, s_{|S|}\}$, wherein we can choose a series over time $\vec{z} \in S^T$. Assume that the transition matrix from a weather system is given by (2)

$$
A = \begin{array}{c} \\ s_0 \\ s_{sun} \\ s_{cloud} \\ s_{rain} \end{array}
\begin{array}{c}
\begin{array}{cccc} s_0 & s_{sun} & s_{cloud} & s_{rain} \end{array} \\
\begin{bmatrix} 0 & 0.4 & 0.5 & 0.1 \\ 0 & 0.5 & 0.2 & 0.3 \\ 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0.1 & 0.7 & 0.2 \end{bmatrix}
\end{array}
$$

Compute the probability for sequence of observation

$$\vec{z} = \{z_1 = s_{sun}, z_2 = s_{cloud}, z_3 = s_{cloud}, z_4 = s_{rain}, z_5 = s_{cloud}\}.$$

3) (5)

A)

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

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

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

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

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

B) Suppose, there are a finite set of models $\mathcal{M} = \{M_1, \ldots, M_d\}$, and you are trying to select one among them, which describes the behavior of your data. Describe various techniques for model selection. (3)

C) State following Markov assumptions: (2)

i) Limited Horizon Assumption

ii) Stationary Process Assumption.

4) Let $|\mathcal{H}| = k$, and m, δ be fixed, the with probability at least $1 - \delta$, we have that (5)

A)

$$\varepsilon(\hat{h}) \leq \left(\underset{h\in\mathcal{H}}{min}\ \varepsilon(h)\right) + 2\sqrt{\frac{1}{2m}\log\frac{2k}{\delta}}$$

, which quantifies bias/variance tradeoff in model selection. Starting with uniform convergence, derive the above result.

B) Show that PCA corresponds to variance maximization. (3)

C) Write k-means clustering algorithm. (2)

5) (5)

A)

Suppose you are given a training set $\{x^{(1)}, \ldots, x^{(m)}\}$. You are required model the data by specifying a joint distribution $p(x^{(i)}, z^{(i)}) = p(x^{(i)}|z^{(i)})p(z^{(i)})$, where $z^{(i)} \sim \text{Multinomial}(\phi)$, $\phi_j \geq 0, \sum_{j=1}^{k} = 1$, and $x^{(i)}|z^{(i)} = j \sim \mathcal{N}(\mu_j, \Sigma_j)$. The parameter ϕ_j gives $p(z^{(i)} = j)$. Your model assumes that each $x^{(i)}$ is drawn from one of k Gaussians depending on $z^{(i)}$. This is called mixture of Gaussians model. The parameters of the model are ϕ, μ and Σ. To estimate the model parameter use the likelihood of your data which is given by

$$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log \sum_{z^i=1}^{k} p(x^{(i)}|z^{(i)}; \mu, \Sigma) p(z^{(i)}, \phi).$$

Use EM algorithm to estimate ϕ and μ. The EM algorithm is given by
Repeat until convergence{
(E-step) For each i, set

$$Q_i(z^{(i)}) := p(z^{(i)}|x^{(i)}; \theta)$$

(M-step) Set

$$\theta := \underset{\theta}{argmax} \sum_{i} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}.$$

}

B) Write the algorithms for value iteration and policy iteration. (3)

C) Suppose $x, z \in \mathbb{R}^2$, and consider $K(x, z) = (x^T z)^2$. You know that $K(x, z) = \phi(x)^T \phi(z)$. (2)
Write feature map $\phi(x)$ for the given kernel.

-----End-----