# CS5590: Exam - 2.

28-Oct-2019 (1.5 hours)

ROLL NO.________________

**Note:** Fill in the blanks/boxes appropriately such that the respective statements become true. While filling the blanks/boxes strictly follow the instructions in the respective question appearing immediately after/before the blank/box. You are free to use any standard mathematical symbols like $\pi, e, \Sigma, ||\cdot||, \log, \max$ etc. Answers that are not simplified enough, (correct) answers in wrong format, illegible writings, and those outside the blanks/boxes, will be ignored by the evaluator. Please attempt the problems in rough sheets first and prepare answers for all the blanks/boxes in rough. Then fair copy them in this sheet while respecting the boundaries of the blanks/boxes.

1. Consider the following dataset $\mathcal{D} = \{(0,1,1,-1),(0,2,1,-2),(1,0,-1,1),(2,0,-1,2)\}$. Let us call the first two variables (in the quadraples) as $x = (x_1, x_2)$, the next as $y_1$ and the last as $y_2$. Let us assume $y_1$ is discrete taking only two values and $x_1, x_2, y_2$ are real. For each of the following paradigms provide name of atleast one model and answer the related questions. Cleverly make your choices for models, such that the answers become simple to write.

**Parametric Discriminative Model for predicting $y_1$ at given $x$:**

Name of one example model in this paradigm:

Logistic Regression.

[0.5 Mark]

In this named model:

(a) The parametric form of the function being modeled is:

$$P_w(y_1|x) = e^{w_{y_1}^T x} / \left(e^{w_1^T x} + e^{w_{-1}^T x}\right)$$

Fill this blank with an equation involving $x, y_1$ and model specific terms. LHS must be the value of the function(s) being modeled, and the RHS, the exact parametric form.

Does not exist
diverges
N/A etc.
are also OK

(b) The estimated parameters with the ~~above~~ MCLE algorithm using the data $\mathcal{D}$ are: $\omega = \begin{bmatrix} -\infty \\ \infty \end{bmatrix}$. [1 Mark]

(c) The estimated posterior likelihood of $y_1 = 1$ given $x = (0, 3)$ is: 1. [0.5 Mark]

**Parametric Discriminative Model for predicting $y_2$ at given $x$:** [0.5 Mark]
Name of one example model in this paradigm:
Linear Regression.

In this named model: [0.5 Mark]

(a) The parametric form of the function being modeled is:

$$P_\omega(y_2/x) = e^{-\frac{1}{2}(y_2 - \omega^T x)^2} / \sqrt{2\pi}$$

may include $\sigma^2$ as hyperparameter ~~of parameters~~. etc.

Fill this blank with an equation involving $x, y_2$ and model specific terms. LHS must be the value of the function(s) being modeled, and the RHS, the exact parametric form.

(b) The estimated parameters with the ~~above~~ MCLE algorithm using the data $\mathcal{D}$ are: $\omega = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. [1 Mark]

$(XX^T)^{-1}Xy$

formulae also OK

$x^T(XX^T)^{-1}Xy$
1.5 marks

(c) The estimated mean of posterior likelihood of $y_2$ ~~$=1$~~ given $x = (5, 3)$ is: 2. [1 Mark]

**Parametric Generative Model for predicting $y_1$ at given $x$:** Name of one example model in this paradigm: [1 Mark]
GDA with diagonal covariances.

full covariances is also OK.

In this named model: [0.5 Mark]

(a) The parametric form of the function(s) being modeled is/are:

$$p_\theta(y_1) = \theta_{y_1}, \quad p(x/y_1) = \frac{e^{-\frac{1}{2}(x-\mu_{y_1})^T \begin{bmatrix} \sigma_{y_1}^2 & 0 \\ 0 & \bar{\sigma}_{y_1}^2 \end{bmatrix}^{-1} (x-\mu_{y_1})}}{2\pi\sqrt{\sigma_{y_1}^2 \bar{\sigma}_{y_1}^2}}$$

Fill this blank with an equation involving $x, y_1$ and model specific terms. LHS must be the value of the function(s) being modeled, and the RHS, the exact parametric form.

[1 Mark]

(b) The estimated parameters with the ~~above~~ MM algorithm using the data $\mathcal{D}$ are:

$$\theta_1 = \theta_{-1} = \frac{1}{2}, \quad \mu_1 = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}, \begin{pmatrix} \sigma_{y_1}^2 \\ \bar{\sigma}_{y_1}^2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0.25 \end{bmatrix}, \mu_{-1} = \begin{bmatrix} 1.5 \\ 0 \end{bmatrix}, \begin{pmatrix} \sigma_{-1}^2 \\ \bar{\sigma}_{-1}^2 \end{pmatrix} = \begin{bmatrix} 0.25 \\ 0 \end{bmatrix}$$

.

[1 Mark]

(c) The estimated posterior likelihood of $y_1 = 1$ given $x = (\pi^{0.5}, 2.5)$ is: $e^{-2} / [e^{-2} + e^{-2(\pi^{0.5} - 1.5)^2}]$

$\boxed{\text{Typo}}$ of 0.5 would have given ½ as neat answer!

[1 Mark]

**Parametric Generative Model for predicting $y_2$ at given $x_1$:** Name of one example model in this paradigm: Gaussian.

[0.5 Mark]

In this named model:

(a) The parametric form of the function being modeled is:

$$p_{\mu,\Sigma}(x_1, y_2) \equiv p_{\mu,\Sigma}(z) = \frac{e^{-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)}}{2\pi\, |\Sigma|^{1/2}}$$

Fill this blank with an equation involving $x_1, y_2$ and model specific terms. LHS must be the value of the function(s) being modeled, and the RHS, the exact parametric form.

[1 Mark]

(b) The estimated parameters with the ~~above~~ MM algorithm using the data $\mathcal{D}$ are:

$$\mu = \begin{bmatrix} 0.75 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 11/16 & 5/4 \\ 5/4 & 5/2 \end{bmatrix} \longrightarrow \begin{bmatrix} 11/12 & 5/3 \\ 5/3 & 10/3 \end{bmatrix} \text{ also OK.}$$

formula above OK
$\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1)$
1.5 marks

[1 Mark]

(c) The estimated mean of posterior likelihood of $y_2$ given $x_1 = 5$ is: $85/11$.

[2 Marks]

2. Consider a model in the exponential family whose sufficient statistics is given by $\phi(x) = x,\ x \in [0, \infty)$. Let $w \in \mathcal{W} \subset \mathbb{R}$ be the parameters for this model. The largest set $\mathcal{W}$ such that for each $w \in \mathcal{W}$, the corresponding function is indeed a valid likelihood is given by: $\mathcal{W} =$ $(-\infty, 0)$. Your expression in the previous blank must involve constant numerals only and must not involve any unknowns.

[2 Marks]

For this model, the simplified expression for the partition function is $Z(w) = -1/\omega$. Your expression in the previous blank must NOT explicitly involve integrals and must not involve any unknowns.

[2 Marks]

Let $\mathcal{D} = \{5, 6, 4\}$ be the training data. The simple equality condition that needs to be satisfied by the optimal parameter $w^*$ for being an MLE solution over $\mathcal{D}$ is: $E_{\omega}[X] = 5$. The LHS in the previous blank must be an expression involving an expectation, and the RHS must be a number.

[1 Mark]

When simplified this equality is: $\omega^* =$ $-1/5$. The LHS in the previous blank must be an expression involving $w^*$ alone, without explicitly involving expectations, and the RHS must be a number.

[1 Mark]