姓名：

学号：

学院：

上海科技大学

2022-2023 学年第 1 学期 期末考试卷

开课单位：生物医学工程学院

授课教师：李远宁

考试科目：神经信号处理与数据分析

课程代码：BME2111

考生须知：

1．请严格遵守考场纪律，禁止任何形式的作弊行为。

2．参加闭卷考试的考生，除携带必要考试用具外，书籍、笔记、掌上电脑和其他电子设备等物品一律按要求放在指定位置。

3．参加开卷考试的考生，可以携带教师指定的材料独立完成考试，但不准相互讨论，不准交换材料。

考试成绩录入表：

题目	1	2	3	**总分**
计分				
复核				

评卷人签名： **复核人签名：**

日期： **日期：**

1. Multiple choices. For each question, select only ONE most appropriate answer. (Total points 40pts = 4pts x 10 + 2 bonus problems)

(1) Which scientist(s) first described the structure of neuron and proposed the neuron doctrine: individual neurons are the elementary signaling elements of the nervous system in the late 19th century?

A. Ramon y Cajal B. Charles Darwin

C. Hodgkin and Huxley D. Hubel and Wiesel

(2) Which of the following is NOT a characteristic of an action potential?

A. All-or-none response B. Self-regeneration

C. Variable strength D. Refractory period

(3) Which of the following ion channels is responsible for the depolarization phase of the action potential?

A. Potassium (K) channels B. Sodium (Na) channels

C. Calcium (Ca) channels D. Chloride (Cl) channels

(4) Which of the following is NOT a step in the process of synaptic transmission?

A. Neurotransmitter release B. Neurotransmitter binding to receptors

C. Depolarization of the postsynaptic neuron D. Repolarization of the presynaptic neuron

(5) What is a probabilistic generative classifier?

A. A machine learning algorithm that uses probability distributions to model the relationships between features and labels in the data.

B. A machine learning algorithm that uses decision trees to model the relationships between features and labels in the data.

C. A machine learning algorithm that uses support vector machines to model the relationships between features and labels in the data.

D. A machine learning algorithm that uses neural networks to model the relationships between features and labels in the data.

(6) Which of the followings is NOT a common feature used in spike sorting algorithms?

A. Amplitude B. Width C. Shape D. Color

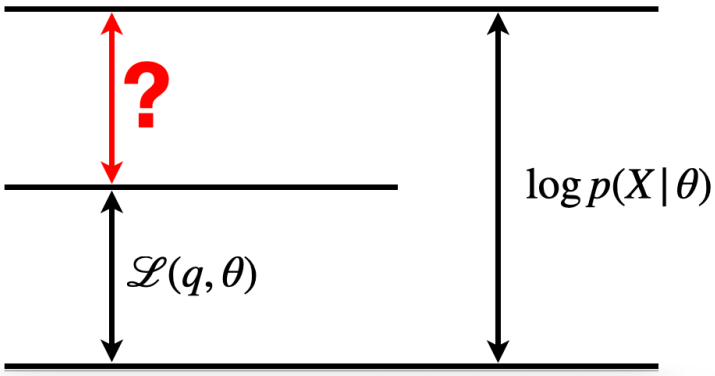
(7) The following figure is a demonstration of the data likelihood decomposition in the EM algorithm. What is the question mark "?" part representing?

A. The prior likelihood $q(Z)$.

B. The posterior likelihood $p(Z|X,\theta)$.

C. The Kullback-Leibler divergence $KL(q||p)$.

D. The joint likelihood $p(X,Z|\theta)$



(8) What is the primary goal of dimensionality reduction?
A. To increase the number of features in a dataset while retaining as much information as possible.
B. To reduce the number of features in a dataset while retaining as much information as possible.
C. To eliminate irrelevant features from a dataset.
D. To eliminate redundant features from a dataset.

(9) What is the advantage of probabilistic PCA (PPCA) over conventional PCA?
A. PPCA assigns probabilities to data, so we can select the dimensionality of low dimensional space and compare to other models using cross-validated likelihoods.
B. PPCA has an explicit noise model, so it is able to more effectively denoise data than PCA.
C. If data dimensionality is large, diagonalization in conventional PCA is costly. If we only need top eigenvectors, we can compute them more efficiently using PPCA.
D. All of the above.

(10) Which of the following is NOT a common method of dimensionality reduction?
A. Principal component analysis (PCA)
B. Linear discriminant analysis (LDA)
C. Logistic regression (LR)
D. Independent component analysis (ICA)

(11) Which of the followings does not have closed form solutions?
A. Principal component analysis (PCA)
B. Linear discriminant analysis (LDA)
C. Linear regression (OLS)
D. Logistic regression (LR)

(12) Which of the following is NOT true regarding Kalman filters?
A. It assumes linear dynamic updates between consecutive states.
B. It is time invariant, i.e. the system parameters do not vary over time.
C. It is noise-free.
D. It can be used for continuous brain-computer interface decoding.

2. Briefly answer the questions. (Total points 60pts = 6pts x 10, +2 bonus questions)

(1) When and why do we need spike sorting?

(2) List three situations where we can use dimensionality reduction.

(3) What is the objective of PCA? What does it try to maximize/minimize?

(4) Do we need to ensure that the data follows Gaussian Distribution when we use PCA? Why?

(5) Why do we need to use Gram trick for PCA? What is the time complexity of PCA using Gram trick given D-dimension data with N samples?

(6) For a prediction model $Y = f(X) + \epsilon$, where ϵ has mean 0, variance σ^2, and is independent to X. Describe what exactly are we trying to evaluate for model assessment and model selection? (You can directly write down the mathematical expression)

(7) The term in (6) can be further decomposed into three parts, what are they?

(8) When we select cross-validation models, we do not always select the model with the minimum CV error, why? How do we select it?

(9) Suppose $\boldsymbol{x} \in \mathbb{R}^D$ is high-dimensional observed data, $z \in \mathbb{R}^M$ is low-dimensional latent variable. Please draw the graphical model of the probabilistic PCA (PPCA) and write down what is $P(\boldsymbol{z})$ and $P(\boldsymbol{x}|\boldsymbol{z})$.

(10) Write down the joint likelihood $P(\boldsymbol{x}, \boldsymbol{z}|\theta)$ of PPCA, what exactly are the model parameters θ?

(11) Suppose $\boldsymbol{x_t} \in \mathbb{R}^D, t = 1, \dots, T$ is high-dimensional observed data, $z \in \mathbb{R}^M, t = 1, \dots, T$ is low-dimensional latent variable. Please draw the graphical model of the linear dynamical system (LDS) and write down what is $P(\boldsymbol{z_t}|\boldsymbol{z_{t-1}})$ and $P(\boldsymbol{x_t}|\boldsymbol{z_t})$.

(12) Write down the joint likelihood $P(\boldsymbol{x_1}, \dots, \boldsymbol{x_T}, \boldsymbol{z_1}, \dots, \boldsymbol{z_T}|\theta)$ of the linear dynamical system (i.e. Kalman filter), what exactly are the model parameters θ?

3. Bonus questions. (Total points 20pts)

Suppose $\boldsymbol{x} \in \mathbb{R}^D$ is observed variable, $\boldsymbol{z} \in \mathbb{R}^M$ is latent variable, $\boldsymbol{\epsilon} \in \mathbb{R}^M$ is multivariate normal random noise. Consider the following latent variable model:

$$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}), \qquad \boldsymbol{\mu} \in \mathbb{R}^D, \mathbf{V} \in \mathbb{R}^{D \times D}$$

$$\boldsymbol{x} = \mathbf{W}\boldsymbol{z} + \boldsymbol{\mu_0}$$

(1) What is the marginal distribution of $\boldsymbol{x}$ (not given $\boldsymbol{z}$)?

(2) If you estimate the parameters $\{\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\mu_0}, \mathbf{W}\}$ using maximum likelihood, how should your converged log-likelihood compare to that if you constrain $\boldsymbol{\mu} = 0, \mathbf{V} = \mathbf{I}_{D \times D}$?