SAMPLE QUESTIONS

PATTERN RECOGNITION (MINOR 2: 23-March-2019)

MLE

1. Derive Maximum Likelihood Estimates (MLE) of unknown mean (μ) and variance (σ^2) when the class conditional density is modeled as Univariate Gaussian density
2. Prove that MLE mean estimator is unbiased.
3. Prove that MLE variance estimator is biased. How can we correct this? Can you show that the corrected estimate is now unbiased?
4. Derive Maximum Likelihood Estimate (MLE) of unknown Mean (μ) but known variance (σ^2) when the class conditional density is modeled as Multivariate Gaussian density.
5. Write down the expression for Maximum Likelihood Estimates (MLE) of unknown Mean (μ) and variance (σ^2) when the class conditional density is modeled as Multivariate Gaussian density.

MAP

6. Derive Maximum A Posteriori Estimate (MAP) of unknown Mean (μ) but known variance (σ^2) when the class conditional density is modeled as Univariate Gaussian density and prior on the parameters also follows Gaussian density.
7. Discuss the case of Maximum A Posteriori Estimate (MAP) of Mean is a convex combination of θ_{MLE} and prior value of the mean (μ_0).

BPE

8. What are Frequentist versus Bayesian approaches and give examples?
9. Derive the Bayesian Parameter Estimate (BPE) of the unknown mean (μ) but known variance (σ^2) of a class conditional density that follows Univariate Gaussian Distribution with a conjugate prior modeled also as Gaussian density.
10. Write down the expression for the above case with Multivariate Gaussian density.
11. Write a short note on Gamma Function
12. Discuss a case when the conjugate prior is not Gaussian Density (for example, BPE Univariate with known mean (μ) but unknown variance (σ^2) , where the conjugate prior is *Gamma Density* function although the class conditional density follows Gaussian Density).
13. State how Maximum Likelihood Estimate (MLE) and Maximum A Posteriori (MAP) estimates are different in their problem setup.
14. Compare and contrast MLE and MAP estimates. How these estimates are related to each other and when is the MAP estimate equivalent to MLE?
15. State how Maximum Likelihood Estimate (MLE) and Bayesian Parameter estimates (BPE) are different in their problem setup. Discuss their strengths and weaknesses.

16.

BPE estimate equations for μ_n and σ_n are as given below. From these equations show that the estimate for mean is a convex combination of Maximum Likelihood Estimate (MLE) and prior. Also comment on the nature of the estimates for mean and variance with respect to sample size.

$$\mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\right)\hat{\mu}_n + \left(\frac{\sigma^2}{n\sigma_0^2+\sigma^2}\right)\mu_0; \text{ where } \hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\sigma_n^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}$$

PCA and FLD

17. What are the `Data Matrix' and `Scatter Matrix' used in principal component analysis (PCA) and how are they related? What is the essential difference between the projections estimated by PCA and Fisher Linear Discriminant (FLD)? Give an example (show graphically) where PCA and FLD give different projections for the same data set.
18. Do Principal Component Analysis (PCA) on the Data matrix X comprising 8 data points in 2-D. Plot the mean-centred data points and draw the principal components (PCs) on the same plot. Comment on how the PCs "represent" the data better. [Hint: Compute the mean vector, scatter matrix, and then find the eigen values and eigen vectors corresponding to the characteristic equation].

$$X = \begin{bmatrix} 1 & 3 & 3 & 5 & 5 & 6 & 8 & 9 \\ 2 & 3 & 5 & 4 & 6 & 5 & 7 & 8 \end{bmatrix}$$

19. Derive the expressions for PCA.
20. Derive the expressions for 2-class FDA.
21. What is the essential difference between Ordinary PCA and Dual PCA? When does one prefer to formulate Dual PCA?
22. What is the motivation for Kernel PCA? Can you sketch the nature of the dataset that Kernel PCA can deal with but not Ordinary PCA or Dual PCA?
23. What is Singular Value Decomposition (SVD)?
24. The following is the Algorithm for Direct PCA (Make appropriate assumptions here about the dimensionality of various matrices. X is $n \times t$, where n is the dimensionality of inputs and t is the number of data samples). Write down the corresponding Algorithm for **Dual PCA**, stating clearly all the relevant assumptions.

Direct PCA Algorithm

Recover basis:	$XX^T = \sum_{i=1}^{t} x_i x_i^T$ and let U = eigenvectors of XX^T corresp. to top d eigenvalues.
Encode training data:	$Y = U^T X$ where Y is a $d \times t$ matrix of encodings of the original data.
Reconstruct training data:	$\hat{X} = UY = UU^T X$.
Encode test example:	$y = U^T x$ where y is a d-dimensional encoding of x.
Reconstruct test example:	$\hat{x} = Uy = UU^T x$.