Select language
English (US)
powered by
PYXIS KIT

Submission Deadline : 21/2/2021 20:00

1 Hours Exam

Q.1 Consider three models M_1, M_2 and M_3 learned from the same data D. Which of the following is/are reasonable cues to select the best model?

You score 2.0 flag score 0

- Select the model that has the largest likelihood (pmeg of the D model) at its MAP location
- Select the model chosen posterior distribution has the smallest variance
- Select the model with the largest marginal likelihood
- Select the model with the largest posterior probability

Q.2 Suppose you have latent variable Bernoulli regression model. Its posterior and posterior predictive distribution. What can happen to the shapes of the equal probability contours of the posterior predictive distribution as the number of observations N increases? Hint: Look at the number of training datum which becomes one or zero. Again, Bernoulli, but one can using not more than 3-4 sentences or 40-words (max, one scenario is fine/ needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.3 Consider a Gaussian mixture model $p(x|\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$, and a Dirichlet prior $p(\pi) = \mathrm{Dir}(\pi|\alpha_0)$. Let $\alpha_0 = \frac{1}{K}\mathbf{1}_K$, and $\alpha_0 = \mathbf{1}_K$ (a vector of ones). Assume a very clean Gaussian (or an exact of the single) $p_k(\mu_k)$ and $p_k(\Sigma_k)$ and π is a vector of zero except one entry is one. Now, we add a new datum x_{N+1} and $x_{N+1} = x_1$ (or any other datum). Which of the following is/are true? Support Vector? Gaussian regression model (GPR)? If yes, Y/N? If no, Y/N? Please do brief in answering.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.4 Which of the following quantities can be obtained given a posterior distribution $p(\theta|X)$?

You score 2.0 flag score 0

- Probability that θ is less than some value θ_0
- The mint estimate of θ (MAE)
- Probability that θ is lies over some range (a,b)
- 95% estimate of θ (MAE)

Q.5 Which of the following priors have equation to zero sparse weight vectors?

You score 2.0 flag score 0

- Gaussian prior with spherical covariance
- Gaussian prior with diagonal covariance
- Spike and slab prior
- Laplace prior

Q.6 In the generalized linear model (GLM) is general-in model? Briefly justify your answer using not more than 3-4 sentences or 40-words.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.7 Which of the following is/are true about a generalized linear model (GLM)? In general?

You score 2.0 flag score 0

- Posterior is Gaussian.
- Posterior predictive is Gaussian.
- MLE has closed form solution
- MAE has a unique solution

Q.8 Suppose you have a prior with probability of heads being $\theta \in [0, 1]$, which is assumed to be a **Beta**(α, β) prior. The coin is assumed independent. Observed and which heads less than 0.5 (it means but you do not know the exact number of heads, but it is known that the number of heads is less than 0.5). Which of the following is/are true? Hint: Look at the MAP estimate. Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is fine/ needed). (Hint: Write the posterior of the normalization constant).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.9 A latent linear Gaussian prior is equivalent to using L2 regularization on the weight vector in Gaussian prior (or used to impose different amounts of regularization on different dimensions). Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is fine/ needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.10 Consider a generalized model for multi-class classification and is a class to learn the MLE and MAP estimate. Now, the class assumed different prior for each class. Which of the following could be the pros and advantage of using the GMP approach over the MLE approach? Answer: Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.11 What is the advantage of selecting the "best" hyperparameter values using an MLE approach as compared to using cross-validation? (Your answer should not use more than 3-4 sentences or 40-words).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.12 MFA latent representation is general-in a good (dec) assumption the posterior distribution of a generalized linear model (GLM)? Ignore any computational cost related issues. Briefly justify your answer using not more than 3-4 sentences or 40-words.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.13 Which of these is more closer against overfitting: (1) Plug in point estimate (dec) (2) Posterior predictive distribution? Briefly justify your answer using not more than 3-4 sentences or 40-words.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.14 Briefly state, (1) the marginal likelihood of a model can also be seen as a special case of the posterior predictive distribution, Answer: Briefly using not more than 3-4 sentences or 40-words (one scenario is needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.15 Consider the probabilistic linear regression model with Gaussian likelihood and Gaussian prior on weight vector $\mathbf{w}$ and hyperparameters α and β. Which of the following is/are true? Hint: Look at the MAP estimate. Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.16 Assume you have K candidate models (assume a uniform prior). Which of the following is/are true? Hint: Look at the MAP estimate. Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is needed).

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.17 Which of the two approaches to super-linear learning - general-in model and discriminative model - would usually require a larger number of parameters to be estimated? Briefly justify your answer using not more than 3-4 sentences or 40-words (max, one scenario is needed). Hint: Look at the number of parameters.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.18 Assume we have $\mathbf{B}$ (latent variable) prior where the prior is $p(\mathbf{B}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{b}_k|\mathbf{0}, \mathbf{I})$ (assume $\mathbf{I}$ is identity matrix). Assume we have a Gaussian prior on $\mathbf{W}$ $p(\mathbf{w}_k|\mathbf{B}) = \mathcal{N}(\mathbf{w}_k|\mathbf{B}\mathbf{b}_k, \mathbf{I})$ (assume $\mathbf{w}_k$ and $\mathbf{b}_k$ are vectors, $\mathbf{B}$ is a matrix). Assume $\mathbf{I}$ is a matrix of ones.

(a) Can (in the) MAP estimate of $\mathbf{w}$ (you may directly write the answer if you know it) and show (or write) the posterior predictive distribution $p(\mathbf{w}_{K+1}|\mathbf{w}_1, \ldots, \mathbf{w}_K)$ (the same as seen in the observations - you don't need to derive it analytically).

(b) What happens to the MAP estimate of $\mathbf{w}$ when the Gaussian prior is centered $\mathbf{b}_k^T$ becomes very small? Briefly justify your answer.

(c) What happens to the MAP estimate of $\mathbf{w}$ when the Gaussian prior is centered $\mathbf{b}_k^T$ becomes very large? Briefly justify your answer.

You score 4.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.19 Which of the following is/are true about Monte Carlo approximation in the context of Bayesian inference?

You score 2.0 flag score 0

- It can be used to approximate/ compute the posterior predictive distribution
- It can be used to approximate/ compute the marginal likelihood
- It used for probabilistic linear regression assuming the hyperparameters α, β are given as input for the posterior predictive distribution
- It the posterior is Gaussian, we do not need to use it.

Q.20 Consider (1-D) observations $\{x_n\}_{n=1}^{N}$, a prior distribution $p(\mu|\tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(\mu-\mu_0)^2}$ where the precision τ is a fixed constant. Assume $p(x_n|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x_n-\mu)^2}$. Assume $\mu_0 = 0$ (a constant). Assume $\tau = 1$ (a constant). Which of the following is/are true? Hint: Look at the posterior distribution $p(\mu|x_1, \ldots, x_N)$ (the posterior is Gaussian). Also write down (just the posterior is needed) τ is a constant. Write an expression for the posterior (write it in terms of τ). Briefly state why?

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.21 The general form of a Gaussian Density Function (pdf) Bernoulli is defined as: $\mathbf{Gaussian}_{\mathbf{x}}(\mathbf{x}, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}$. Here μ and Σ are the mean and covariance parameters respectively and $\mathbf{Gaussian}$ the general function, and the lower case $\mathbf{x}$ is a vector. Suppose that both under the shape and scale parameters are α and β $\mathbf{Gamma}_{\alpha,\beta}(\theta)$.

(a) Approximate the gamma distribution by a Gaussian using Laplace approximation.

(b) Now (dec) the gamma approximation complete with approximating $\mathbf{Gaussian}_{\mathbf{x}}(\mathbf{x}, \mu, \Sigma)$ (approximation of the gamma distribution). Write down the expression for the mean and covariance of the actual gamma distribution $\mathbf{Gaussian}_{\mathbf{x}}(\mathbf{x}, \mu, \Sigma)$. Under what condition would these be the same as the mean and the mode of the gamma.

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

Q.22 Consider the following "toy-deep" model for binary classification.

(1) Draw a latent variable vector and bias, $\mathbf{w}_0 \sim \mathbf{Gaussian}(\mathbf{0}, \lambda_0 \mathbf{I})$ (assume the Gaussian's was parameter Σ_0 itself is a function $\mathbf{Gaussian}(\mathbf{x}, \mathbf{w}_0, \Sigma)$ where $\mathbf{x}$ is another fixed vector, $\mathbf{w}_0$ is a vector, Σ is a matrix). Assume $\mathbf{I}$ is a matrix of ones. Assume λ_0 is a constant. Assume $\mathbf{Gaussian}(\mathbf{x}, \mathbf{w}_0, \Sigma) = \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{w}_0)^T \Sigma^{-1}(\mathbf{x}-\mathbf{w}_0)}$. Assume $\mathbf{x}$ is the same as previous (fixed) vector and have the gaussian density function is parameterized by the shape scale parameter (which is $\mathbf{Gaussian}(\mathbf{x}, \mathbf{w}_0, \Sigma)$) and $\alpha = 0.5$).

(2) Set the bias $\mathbf{b}_0 = 1.0$ (dec). (3) When (dec) $\mathbf{b}_0 = 0$.

For the above model:

(a) Derive the posterior $p(\mathbf{w}_0, \mathbf{b}_0|\text{conditioned on } \mathbf{x})$ and $\mathbf{w}$ (or θ (particular), $\mathbf{b}_0$ (scalar), $\mathbf{x}$ (scalar) and derive the expression for $p(\mathbf{y}_0 = 1|\mathbf{x}, \mathbf{w}, \mathbf{b}_0)$ (dec). Hint: You only find (write) the expression for $p(\mathbf{y}_0 = 1|\mathbf{x})$ and their integrals are $\mathbf{w}$, $\mathbf{b}_0$.

(b) Briefly explain why the expression $p(\mathbf{y}_0 = 1|\mathbf{x}_0, \mathbf{w}, \mathbf{b}_0)$ makes this model equivalent to logistic regression model? (Hint: Look at the sigmoid).

(c) Assuming a Gaussian prior on $\mathbf{w}$ can possible (or one $\mathbf{w}$ be obtained in closed form for this model?

You score 2.0 flag score 0

This is a response or text question. You can enter a response in your space and save.

UPLOAD A FILE

or

$$\mathbf{x}_{N+1}(\mathbf{x}) = \mathbf{x}_1 \mathbf{I}^T \mathbf{x}_1^T \mathbf{x}_1^T \mathbf{x}_1$$

Hint: $p(\mathbf{x}_0 \mathbf{b}_0 = \mathbf{x}_1 \mathbf{x}_0 \mathbf{x}_0 \mathbf{x}_0