CS698X: Topics In Probabilistic Modeling And Inference
Users Online : 5

Quiz 2

Q.1 Consider a Bayesian matrix factorization model where the matrix entries are binary. Assume Gaussian priors on the user and item parameters (similar to what we discussed in the lecture). Which known probabilistic modeling problem will learning each conditional posterior (CP) reduce to in this case? Can we obtain each CP in closed form? If not, suggest a possible approximation for each CP such that we can derive a Gibbs sampler identical to the case discussed in the lecture when the matrix entries were real-valued. NOTE: You do not need to write any equations; just answer the question briefly in plain text.

Max. score: 5; Neg. score: 0; Your score: 3

Your answer:

1. This will reduce to logistic regression as likelihood is bernoulli(sigmoid(u_i^T,v_j))

2. No we can't obtain CP in closed form

3. we can use monte carlo approximation for approx CP

Feedback:

Monte Carlo approximation does not approximate a posterior or conditional posterior. 3/5 marks given

---

Q.2 Both Gaussian Process (GP) as well as Relevance Vector Machine (RVM) are methods for learning nonlinear predictive models and use kernels. Will they be equally fast/slow at test time? Briefly justify your answer (in the textbox below, preferably using plaintext only; no file uploads will be considered).

Max. score: 3; Neg. score: 0; Your score: 3

Your answer:

No guassin process will be slow as RVM contains less number of support vectors to calculate on. where as GP has to consider all the data points

Feedback:

Q.3 Consider a Bayesian linear regression model with an unknown weight vector as well as unknown hyperparameters of the Gaussian prior and Gaussian likelihood. Which of the following are true?

CS698X: Topics In Probabilistic Modeling And Inference
Users Online : 5

on it) of the Gaussian prior will have closed form expression

- ✔ ☐ The conditional posterior for the precision hyperparameter (assuming a gamma prior on it) of the Gaussian likelihood will have closed form expression
- ✔ ☐ The conditional posterior on the weight vector will have closed form expression
- ☐ Only the conditional posterior of the weight vector will have closed form expression but conditional posteriors of **neither** of the hyperparameters (assuming gamma priors on each) have closed form expressions

---

Q.4 Can we use a kernelized SVM for learning the underlying function when doing Bayesian Optimization? Briefly justify your answer (in the textbox below, preferably using plaintext only; no file uploads will be considered).

Max. score: 3; Neg. score: 0; Your score: 3

Your answer:

No it can't be used in kernazed svm as it requires some model which can give the estimate/varaince in the prediction.

Feedback:

Q.5 For a training set with N labeled example, a GP will be equivalent to a single hidden layer neural network with N hidden units.

Max. score: 1; Neg. score: 0; Your score: 1

- ✔ ◉ false
- ○ true

Q.6 Considering using Bayesian Optimization to minimize some function. The lower confidence bound (LCB) acquisition function is defined as $A_{LCB}(x_{new}) = \mu(x_{new}) - \kappa\sigma(x_{new})$ where $\mu(x_{new})$ and $\sigma(x_{new})$ denote the mean and standard deviation of the posterior predictive of x_{new}, and the most useful query point is chosen as the one that **minimizes** this acquisition function. Provide a brief justification as to why the use of such an acquisition function makes sense (answer in the textbox below, preferably using plaintext only; no file uploads will be considered).

Max. score: 3; Neg. score: 0; Your score: 3

Your answer:

CS698X: Topics In Probabilistic Modeling And Inference
Users Online : 5

Feedback:

Q.7 Which of the following is true about Gaussian Processes (GP)?

Max. score: 2; Neg. score: 0; Your score: 0

- [ ] The hyperparameters can only be learned using MLE-II for GP regression models with Gaussian likelihoods, but not for GP classification or GP GLM models.
- [ ] A GP based classification model would be faster at test time that a kernel SVM model
- ✔ [x] When used for supervised learning problems, they are discriminative models.
- [x] Assuming fixed hyperparameters, the prior as well as the posterior is Gaussian for GP based models

---

Q.8 We can use a kernelized SVM for learning the underlying function when doing Bayesian Optimization.

Max. score: 1; Neg. score: 0; Your score: 0

- ✔ ( ) false
- (•) true

Q.9 For which of the following models (assuming hyperparameters fixed), the posterior predictive can be computed without first computing the posterior distribution?

Max. score: 2; Neg. score: 0; Your score: 0

- [x] GP binary classification with Bernoulli likelihood
- ✔ [ ] Bayesian linear regression with Gaussian likelihood
- ✔ [x] GP regression with Gaussian likelihood
- [ ] Logistic regression with Bernoulli likelihood

Q.10 Entropy or variance of the posterior predictive of the query input is often used as an acquisition function for Active Learning. Will it be a good choice as the acquisition function if we are solving a

CS698X: Topics In Probabilistic Modeling And Inference
Users Online : 5

Your answer:

No its not as BO main aim is to find the optima of the function. The variance acquisition will only sample the next point where it has maximum variance and do not consider the point where its expectation is minimum.

Feedback:

Q.11 A conditional posterior may not always be tractable.

Max. score: 1; Neg. score: 0; Your score: 1

- false
- ✔ true

Q.12 A kernelized model such as Gaussian Process (GP) does not have weights for each feature unlike a linear model so we can't use methods like ℓ_1 or ℓ_0 regularization of the weights. Can you learn (not using a heuristic but through proper learning!) the feature importances when using a GP based model, in order to perform feature selection? If no, briefly state why. If yes, briefly state how (answer in the textbox below, preferably using plaintext only; no file uploads will be considered).

Max. score: 3; Neg. score: 0; Your score: 0

Your answer:

Feedback:

Q.13 Active Learning with maximum entropy criterion is equivalent to selecting query inputs randomly from the unlabeled pool of inputs.

Max. score: 1; Neg. score: 0; Your score: 1

- ✔ false
- true

Q.14 A Gaussian Process regression model with a linear kernel is equivalent to a Bayesian linear regression model (assuming fixed hyperparameters).

CS698X: Topics In Probabilistic Modeling And Inference
Users Online : 5

true

Q.15 Consider a supervised learning model where each response y_n is modeled via some likelihood model $p(y_n|f_n)$ with $f_n = f(x_n)$ and the function f is modeled by a Gaussian Process. Given a test input x_* and f's values at N training examples $\mathbf{f} = [f(x_1), f(x_2), \ldots, f(x_N)]$, which of the following are true {(assume hyperparameters as fixed)?

Max. score: 2; Neg. score: 0; Your score: 2

- The predictive distribution of f's values at the test input, i.e., $p(f_*|f_1, f_2, \ldots, f_N)$ is a Gaussian only if each likelihood $p(y_n|f_n)$ is Gaussian
- The posterior predictive distribution of the actual label of the test input, i.e., $p(y_*|y_1, y_2, \ldots, y_N)$ is always a Gaussian
- ✔ The predictive distribution of f's values at the test input, i.e., $p(f_*|f_1, f_2, \ldots, f_N)$ is always a Gaussian
- The posterior predictive distribution of the actual label of the test input, $p(y_*|y_1, y_2, \ldots, y_N)$ is never a Gaussian

---

Q.16 Which of the following quantities can be used as reasonable criteria to select the most informative query inputs for Active Supervised Learning?

Max. score: 2; Neg. score: 0; Your score: 0

- ✔ Amount of decrease in the posterior's uncertainty upon including the query input
- Mean of the posterior predictive of the label of the query input
- ✔ Uncertainty in the posterior predictive of the label of the query input
- Amount of noise in the query input

Score: 21