

Вопросы к экзамену по курсу “Байесовские методы в анализе данных”

На экзамене можно пользоваться одним листом размера A5 (половина A4), написанным от руки (с двух сторон).

Во всех вопросах нужно уметь расшифровывать обозначения в формулах!

Введение в байесовские методы

1. Запишите формулы условной вероятности и формулу Байеса.
2. В чем состоят отличия оценки параметров в байесовском и частотном подходе?
3. Что такое метод максимального правдоподобия? В чем состоят его недостатки?
4. Запишите определение сопряженного семейства распределений. Приведите два примера сопряженных семейств распределений.
5. Какие распределения необходимо задать, чтобы задать байесовскую модель машинного обучения?
6. Какое распределение необходимо искать на этапе обучения байесовской модели машинного обучения? А на этапе предсказания?
7. Что такое оценка максимума апостериорной вероятности? В чем ее преимущества и недостатки по сравнению с методом максимального правдоподобия и с полным байесовским выводом?

Полный байесовский вывод для линейной регрессии

8. Запишите формулу плотности многомерного нормального распределения. Как интерпретируются параметры этого распределения? Какова мода этого распределения? Как выглядят линии уровня плотности многомерного нормального распределения при диагональной и недиагональной матрице ковариации?
9. Запишите модель байесовской линейной регрессии.
10. Поясните, почему поиск максимума апостериорной вероятности в модели байесовской линейной регрессии соответствует L2-регуляризации модели.
11. Запишите алгоритм обучения модели байесовской линейной регрессии. Каковы гиперпараметры этой модели?
12. По какой формуле будут вычисляться среднее и дисперсия предсказания для нового объекта в модели байесовской линейной регрессии?
13. Для чего на практике может использоваться дисперсия предсказания в байесовской линейной регрессии?
14. Что такое обоснованность модели? Для чего она используется?
15. Запишите модель метода релевантных векторов для задачи регрессии. Каким образом в этой модели производится отбор признаков?

Вариационный вывод и байесовские нейронные сети

16. Запишите формулу KL-дивергенции. Каковы ее свойства?
17. Каким образом в вариационном выводе ищут приближенное апостериорное распределение? Запишите функционал, который оптимизируют в вариационном выводе, и объясните, откуда он получается. (очень важный вопрос!)
18. Выведите вариационную нижнюю оценку $L(q)$ на логарифм правдоподобия $\log p(x)$ для произвольного распределения q на множестве параметров модели.
19. Почему два слагаемых вариационной нижней оценки называют data term и regularizer?
20. По каким величинам оптимизируют вариационную нижнюю оценку в параметрическом вариационном выводе?
21. В чем отличия байесовской нейронной сети от детерминированной нейронной сети? В чем состоят преимущества использования байесовской нейронной сети?
22. Какое распределение ищут на этапе предсказания с помощью байесовской нейронной сети? Как его приближают?
23. За счет какого механизма байесовские нейронные сети достигают высокого качества в оценке неопределенности в предсказании (uncertainty estimation)?
24. Какой функционал оптимизируют при обучении этой модели? Как вычислять/оценивать слагаемые в этом функционале?
25. Что такое трюк репараметризации (reparametrization trick)? Для чего он нужен? Запишите трюк в общем виде и приведите пример.
26. Почему использование дропаута можно интерпретировать как обучение байесовской нейронной сети? В чем преимущества такой интерпретации?
27. Какое априорное и приближенное апостериорное распределения используют в модели Sparse variational dropout? За счет какого механизма достигается разреживание весов нейронной сети в этой модели?
28. Запишите алгоритм прохода вперед при обучении модели Sparse variational dropout. В какой момент и каким образом выполняется обнуление весов? Как выполняется проход вперед на этапе предсказания?

MCMC

29. Каким образом семплирование (метод Монте-Карло) используется в байесовском машинном обучении? Почему можно считать, что метод Монте-Карло заменяет полный байесовский вывод?
30. Что такое rejection sampling и importance sampling? Позволяют ли эти методы семплировать из распределения, известного с точностью до нормировочной константы? Если да, то каким образом?
31. Что такое распределение, инвариантное относительно марковской цепи? Что такое эргодичная марковская цепь? Запишите и докажете достаточное условие инвариантности.

32. Запишите схему Метрополиса-Хастингса семплирования выборки из распределения, известного с точностью до нормировочной константы.
33. В чем отличие схемы Метрополиса от схемы Метрополиса-Хастингса?

Гауссовские процессы

34. Что такое гауссовский процесс? Приведите два примера (функция среднего и ядровая функция, указать, какие функции генерируются соответствующим гауссовским процессом)
35. В чем состоит модель гауссовского процесса для задачи регрессии? Запишите формулу для среднего и дисперсии предсказания на новом объекте в этой модели.
36. В чем состоит обучение гауссовского процесса для задачи регрессии? Какие величины необходимо настраивать?
37. Каковы преимущества и недостатки использования гауссовских процессов на практике?
38. В чем отличие параметрических и непараметрических методов машинного обучения? Приведите по два примера из каждой группы.
39. Какую задачу решает байесовская оптимизация? Запишите алгоритм байесовской оптимизации. Приведите примеры использования байесовской оптимизации.

ЕМ-алгоритм для моделей со скрытыми переменными

40. Приведите два примера моделей со скрытыми переменными.
41. Запишите, какой функционал оптимизируют в ЕМ-алгоритме. Нижней оценкой на какую величину он является? Запишите формулы Е- и М-шагов. Поясните, как выводятся эти формулы.
42. Запишите модель вероятностного метода главных компонент. Что является наблюдаемыми, скрытыми переменными и параметрами в этой модели?
43. Запишите модель смеси гауссиан. Что является наблюдаемыми, скрытыми переменными и параметрами в этой модели?
44. Запишите формулу Е-шага для модели смеси гауссиан. Запишите, какой функционал необходимо минимизировать на М-шаге, и укажите, по каким величинам его необходимо минимизировать.
45. Выведите формулу М-шага для обновления центров гауссиан в модели смеси гауссиан.

Вариационные автокодировщики и репараметризация дискретных распределений

46. Какую задачу решает вариационный автокодировщик?
47. Запишите модель вариационного автокодировщика.
48. Какой функционал оптимизируют при обучении вариационного автокодировщика (итоговый функционал)? Поясните, как интерпретируются слагаемые функционала.
49. Запишите трюк репараметризации для вариационного автокодировщика.
50. Запишите алгоритм обучения вариационного автокодировщика (алгоритм обработки одного мини-батча).
51. Запишите алгоритм генерации новых объектов с помощью вариационного автокодировщика.
52. В чем преимущества и недостатки использования вариационных автокодировщиков по сравнению с Generative adversarial networks (GAN)?
53. Какой функционал оптимизируют при обучении GAN?
54. Для чего используются Gumbel softmax trick и REINFORCE?
55. Запишите алгоритм семплирования с помощью Gumbel softmax trick. На что влияет температура? Какие получатся семплы, если использовать нулевую и очень высокую температуру? Каковы достоинства и недостатки этого метода?
56. Запишите алгоритм семплирования с помощью REINFORCE. Как выводится этот алгоритм? Каковы достоинства и недостатки этого метода?

На экзамене могут быть задачи с такими формулировками (не исчерпывающий список):

- Доказать, что два распределения сопряжены
- Вычислить среднее апостериорного распределения в модели с сопряженными распределениями
- Записать схему МСМС для заданного распределения
- Какой метод байесовского вывода применим к заданной модели?
- Задачи из непроверяемых домашних заданий
- Задачи, которые разбирались на занятиях (см. презентации)