

ПРИЛОЖЕНИЕ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ

Кафедра «Автоматика и информационные технологии в управлении»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине

«Машинное обучение»

Направление подготовки

01.03.02 Прикладная математика и информатика

ОПОП академического бакалавриата

«Прикладная математика и информатика»

Квалификация (степень) выпускника – бакалавр

Формы обучения – очная

Рязань 2024г.

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Использует фундаментальные знания, полученные в области математических наук при решении научных и технических задач в своей профессиональной деятельности
	ОПК-1.2 Использует фундаментальные знания, полученные в области естественных наук при решении научных и технических задач в своей профессиональной деятельности
ОПК-2 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач	ОПК-2.1. Использует существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач
	ОПК-2.2. Адаптирует существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение» обучающийся должен:

знать:

основные принципы и проблематику теории обучения машин, основные современные методы обучения по прецедентам — классификации, кластеризации и регрессии.

уметь:

формализовать постановки прикладных задач анализа данных, использовать методы обучения по прецедентам для решения практических задач, оценивать точность и эффективность полученных решений.

владеть:

основными понятиями теории машинного обучения.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Построение системы машинного перевода
Реализация DQN с памятью (experience replay buffer)
Реализация метода Advantage Actor Critic в среде Atari
Классификация и сегментация изображений
Построение генеративных моделей

перечень вопросов по теорминимуму.

1. Постановка задач обучения с учителем (supervised learning).
2. Задачи обучения без учителя. Назвать хотя бы две.
3. Что означает свойство i.i.d.?
4. Основная идея наивного Байесовского классификатора. В чём его наивность?
5. Запишите формулы для модели линейной регрессии и для среднеквадратичной ошибки.
6. Запишите формулу для одного шага градиентного спуска. Как модифицировать градиентный спуск для очень большой выборки?
7. Что такое правдоподобие, метод максимального правдоподобия? Является ли правдоподобие вероятностью?
8. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?
9. Что такое переобучение и недообучение? Как их можно детектировать?
10. Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?

11. Что такое регуляризация? Чем на практике отличается L1-регуляризация от L2?
12. Учитывается ли коэффициент сдвига w_0 в регуляризаторе? Почему?
13. Почему линейные модели рекомендуется применять к выборке с нормированными значениями признаков?
14. Запишите формулу для линейной модели классификации. Что такое отступ?
15. Что такое точность и полнота? Почему нужно учитывать их вместе?
16. В задаче бинарной классификации доля одного класса составляют 95% выборки. Какие метрики разумно использовать для оценки работы модели? почему?
17. Что такое ROC-AUC? Как построить ROC-кривую?
18. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
19. Идея метода опорных векторов (в случае разделимой выборки).
20. Опишите жадный алгоритм обучения решающего дерева.
21. Почему с помощью решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?
22. Если в лист дерева попали объекты разных классов, то какие предсказания нужно выдавать в этом листе? Почему?
23. Какое предсказание нужно выдавать в листе дерева в задаче регрессии если мы минимизируем MSE? а в случае MAE?
24. Что такое bagging?
25. Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?
26. Как в градиентном бустинге обучаются базовые алгоритмы?
27. Зачем нужен backprop, что такое производная вектора по вектору?
28. Опишите принцип работы свёрточного слоя (CNN).
29. В чем недостатки полносвязных нейронных сетей какая мотивация к использованию свёрточных?
30. Опишите принцип работы базового рекуррентного слоя (RNN).
31. Что такое dropout?
32. Как dropout и batch normalization меняют свое поведение при эксплуатации модели (в режиме inference)?
33. Запишите постановку задачи в методе главных компонент.
34. Как работает метод k-means?

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Задачи обучения по прецедентам. Supervised, unsupervised и semi-supervised обучение. Понятия переобучения и обобщающей способности. Скользящий контроль (cross-validation).
2. Метрические алгоритмы классификации. Обобщённый метрический классификатор, понятие отступа. Метод ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля. Отбор эталонных объектов. алгоритм СТОЛП. Функция конкурентного сходства (FRiS).
3. Построение метрик и отбор признаков. Стандартные метрики. Оценивание качества метрики. Проклятие размерности. Жадный алгоритм отбора признаков.
4. Логические закономерности. Статистический критерий информативности I с $(\varphi, X | I)$: смысл и способы вычисления. Энтропийный критерий информативности — информационный выигрыш $IGain$ с $(\varphi, X | I)$. Многоклассовые варианты критериев. Индекс Gini. Задача перебора конъюнкций. “Градиентный” алгоритм синтеза конъюнкций и его частные случаи: жадный алгоритм, стохастический локальный поиск, стабилизация, редукция.
5. Бинаризация признаков, алгоритм выделения информативных зон. Решающие списки. Решающие деревья: принцип работы. Разбиение пространства объектов на подмножества, выделяемые конъюнкциями терминальных вершин. Алгоритм ID3. Пре-прунинг и пост-прунинг. RandomForest.

6. Линейная классификация. Непрерывные аппроксимации пороговой функции потерь. Метод минимизации аппроксимированного эмпирического риска. SG, SAG. Связь минимизации аппроксимированного эмпирического риска и максимизации совместного правдоподобия данных и модели. Регуляризация (l1, l2, elasticnet). Вероятностный смысл регуляризаторов. Примеры различных функций потерь и классификаторов. Эвристический вывод логистической функции потерь.
7. Метод опорных векторов. Оптимизационная задача с ограничениями в виде неравенств и безусловная. Опорные векторы. Kerneltrick. Оптимизационная задача в S3VM и SVR. SVM и беспризнаковое машинное обучение на примере ядер графов и классификации вершин графа.
8. Задача снижения размерности пространства признаков. Идея метода главных компонент (PCA). Связь PCA и сингулярного разложения матрицы признаков (SVD). Вычисление SVD в пространствах высокой размерности методом стохастического градиента (SG SVD).
9. Многомерная линейная регрессия. Геометрический и аналитический вывод. Регуляризация в задаче регрессии. Непараметрическая регрессия. Формула Надарая-Ватсона. Регрессионные деревья.
10. Байесовская классификация и регрессия. Функционал риска и функционал среднего риска. Оптимальный байесовский классификатор и теорема о минимизации среднего риска. Наивный байесовский классификатор.
11. Восстановление плотности: параметрический и непараметрический подход. Метод Парзенковского окна. Параметрический подход на примере нормального дискриминантного анализа. Линейный дискриминант Фишера.
12. Задача прогнозирования временного ряда, примеры задач. Адаптивные алгоритмы прогнозирования: экспоненциальное сглаживание, модели Брауна, Тейла-Вейджа, Хольта-Винтерса. Преимущества и недостатки адаптивных алгоритмов прогнозирования.
13. Модели ARMA, ARIMA, а также регрессионные методы решения задачи прогнозирования временного ряда. Композиции адаптивных алгоритмов: селекция, композиция, LAMP, агрегирующий алгоритм.
14. Задача кластеризации. Агломеративная и дивизионная кластеризация. Алгоритмы k-Means, k-Means++. Кластеризация с помощью EM-алгоритма (без вывода M-шага). Формула Ланса-Уилльямса.

Пример для экзаменационного билета из трех вопросов:

1. Этапы построения модели: тренировка, валидация, тестирование. Роль каждого этапа. Переобучение.
2. Регуляризация нейронных сетей: Dropout, Batch Normalization. Разница в поведении при обучении и предсказании
3. Логистическая регрессия. Эквивалентность решений, полученных методом максимального правдоподобия и минимизации логистической функции потерь

Критерии оценивания

отлично

10 всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений;

9 систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений;

8 глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, правильное обоснование принятых решений;

хорошо

7 твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;

6 знает материал, грамотно излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;

5 знает основной материал, грамотно излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач неточности;

удовлетворительно

4 фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

3 характер знаний достаточен для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

неудовлетворительно

2 не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет правильно использовать полученные знания при решении типовых практических задач.

1 не знает формулировок основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

За первое задание студент получает от 0 до 4 баллов, за второе и третье – от 0 до 3 баллов за каждое в зависимости от полноты представленного ответа (решения). Количество набранных баллов определяет оценку за экзамен.

При проведении экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету не должен превышать двух астрономических часов.

Во время проведения экзамена обучающиеся могут пользоваться программой дисциплины.

Оператор ЭДО ООО "Компания "Тензор"

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

СОГЛАСОВАНО

ФГБОУ ВО "РГРТУ", РГРТУ, Бабаян Павел Варганович,
Заведующий кафедрой АИТУ

14.08.24 09:44 (MSK)

Простая подпись