

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Физический факультет

УТВЕРЖДЕНО:
декан физического факультета
С.Н. Филимонов

Оценочные материалы по дисциплине

Основы машинного обучения

по направлению подготовки

03.03.02 Физика

Направленность (профиль) подготовки / специализация:
«Фундаментальная и прикладная физика»

Форма обучения
Очная

Квалификация
Бакалавр

Год приема
2025

СОГЛАСОВАНО:
Руководитель ОП
С.Н. Филимонов

Председатель УМК
О.М. Сюсина

Томск – 2025

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

– ОПК-3 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности..

– ПК-1 Способен проводить научные исследования в выбранной области с использованием современных экспериментальных и теоретических методов, а также информационных технологий.

– УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

–ИОПК 3.1 Знает основы программирования и требования информационной безопасности

– ИОПК 3.2 Применяет общее и специализированное программное обеспечение для теоретических расчетов и обработки экспериментальных данных

– ИПК 1.2 Владеет практическими навыками использования современных методов исследования в выбранной области

– ИУК 1.1 Осуществляет поиск информации, необходимой для решения задачи

– ИУК 1.2 Проводит критический анализ различных источников информации (эмпирической, теоретической)

– ИУК 1.3 Выявляет соотношение части и целого, их взаимосвязь, а также взаимоподчиненность элементов системы в ходе решения поставленной задачи

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

– Домашнее задание

– Самостоятельная работа студента на практике

Пример

Домашнее задание:

Задача 1. Написать функцию которая линейно нормализует значения некоторого признака в диапазоне от 0 до 1. Предполагается что признак является столбцом матрицы в виде `np.ndarray`.

Задача 2. Напишите функцию, которая вычисляет значение метрики `recall` для последовательности предсказаний `y_pred` (вектор целых чисел 0 или 1), последовательности правильных значений `y` (булев вектор) при заданном пороге классификации `th`.

Ответы:

Задача 1:

```
def linear_normalize(x):
    delta = x.max() - x.min()
    return (x - x.min()) / delta
```

Задача 2:

```
def recall(y, y_pred, th)
    y_pred_bin = (y_pred > th).astype(int)

    TP = np.sum((y_pred_bin == 1) & (y == 1))
    P = np.sum(y == 1)
    return TP / P
```

Самостоятельная работа на паре:

Задача 1. Используя библиотеку sklearn построить линейную модель на наборе данных и оценить среднеквадратичную ошибку модели. Предполагается, что набор данных состоит из матрицы признаков X, которая задана в виде матрицы (np.ndarray), строки которые соответствуют объектам, а столбцы - признакам, а ответы - в виде вектора-столбца Y (также np.ndarray).

Задача 2. Используя библиотеку sklearn построить линейную модель на наборе данных и оценить значимость признаков. Предполагается, что набор данных состоит из матрицы признаков X, которая задана в виде матрицы (np.ndarray), строки которые соответствуют объектам, а столбцы - признакам, а ответы - в виде вектора-столбца Y (также np.ndarray).

Ответы:

Задача 1:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X, y)

y_pred = model.predict(X)
MSE = ((y_pred - y) ** 2).mean()
print('Среднеквадратичная ошибка =', MSE)
```

Задача 2:

```
from sklearn.linear_model import LinearRegression
import numpy as np

model = LinearRegression()
model.fit(X, y)

abs_coefs = np.abs(model.coef_)
top_features = np.argsort(abs_coefs)[::-1]

print('Признаки в порядке значимости:')

for N in zip(top_features):
    print(f'Признак №{N}; Значение: {model.coef_[N]:.1e}')
```

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Зачет с оценкой проводится в устной форме по билетам. Билет содержит два вопроса: один общий, второй про алгоритмы (или модели)

Теоретические вопросы:

- Постановка задачи машинного обучения.
- Функции потерь. Требования к функциям потерь. Функции потерь для задач классификации.
- Функции потерь. Требования к функциям потерь. Функции потерь для задач регрессии.
- Метрики качества. Требования к метрикам качества. Метрики качества для задач классификации.
- Метрики качества. Требования к метрикам качества. Метрики качества для задач регрессии
- Переобучение. Обнаружение и устранение.
- Кросс-валидация и k-Fold.
- Матрица ошибок. Ошибки первого и второго рода. Производные метрики качества.
- Выбор порога классификация. Кривая точность-полнота и ROC-кривая.\
- Параллельное и последовательное ансамблирование моделей.
- Метод градиентного спуска и метод обратного распространения ошибки.
- Математическая модель искусственного нейрона.

Алгоритмы и модели:

- Метод k-ближайших соседей.
- Метод опорных векторов.
- Непараметрическая регрессия. Формула Надарая-Уотсона.
- Линейная регрессия.
- Логистическая регрессия.
- Линейная регрессия по комбинации базисных функций.
- Решающие деревья.
- Random Forest
- Градиентный бустинг над решающими деревьями
- Полносвязные ИНС
- Свёрточные ИНС
- Автокодировщики

Для получения оценки: **отлично** - студент должен уверенно ответить на оба вопроса из билета, а также на дополнительные вопросы; **хорошо** - студент должен ответить на оба вопроса из билета, но потеряться на дополнительных. Оценка **удовлетворительно** ставится если на один из вопросов билета студент ответить не смог.

Допускаются пробелы в ответах, которые правильно закрываются наводящими вопросами преподавателя.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Тест (ИОПК 3.1)

1. Какая метрика не является подходящей метрикой для задачи регрессии:
 - 1) Точность (accuracy)
 - 2) Средний квадрат отклонения (MSE)
 - 3) Средний модуль отклонения (MAE)

4) Медиана относительного отклонения (MdAPE)

2. Категориальные признаки кодируются при помощи:

- 1) Линейной нормализации
- 2) Экспоненциальной нормализацией
- 3) Заменой уникальных категорий на целые числа
- 4) Заменой столбца на несколько, соответствующих уникальным категориям

Ключи: 1 – 1, 2 – 4

Задачи (ИОПК 3.3)

Задача 1. Сравнить распределение целевых признаков в обучающей выборке (вектор y) и среди откликов обученной модели (вектор y_{pred}).

Задача 2. Написать процедуру разбиения набора данных на обучающую и валидационную подвыборку. Предполагается, что набор данных состоит из матрицы признаков X , которая задана в виде матрицы (`np.ndarray`), строки которые соответствуют объектам, а столбцы - признакам

Ответы:

Задача 1. На одних осях строятся гистограммы распределений y и y_{pred} , при совпадении семейств распределения оцениваются соответствующие параметры (например, для нормального это выборочное среднее и дисперсия)

Задача 2.

```
def train_test_split(X, y, val_percent):
    train_mask = np.random.rand(X.shape[0]) > val_percent

    X_train = X[train_mask]
    y_train = y[train_mask]

    X_val = X[~train_mask]
    y_val = y[~train_mask]

    return (X_train, y_train), (X_val, y_val)
```

Теоретические вопросы: (ИОПК 3.2)

1. *Сформулируйте разницу между задачей регрессии и классификации*

Ответ должен содержать рассуждения о характере целевого признака. Так, для задачи регрессии целевой признак должен принадлежать некоторому непрерывному множеству с определённой на нём метрикой, а для задачи классификации - некоторому дискретному множеству без операции сравнения (больше-меньше) на нём

2. *Допустимые способы заполнения пропущенных значений в обучающей выборке*

Ответ должен содержать перечисления способов: удаление столбца; замена средним, медианой или модой по столбцу; заменой специальным значением, указывающим на пропуск но поддающимся обработкой (учетом) заданной моделью

Информация о разработчиках

Красавин Дмитрий Сергеевич, Томский государственный университет, м.н.с.

Галушина Татьяна Юрьевна, к.ф.-м.н., Томский государственный университет,
доцент