

ML ДПО 1: Классификации

*Обязательный вопрос

1. Электронная почта *

2. Сопоставьте задачу и тип классификации:

* 3 балла

Задача 1: разделите тексты на позитивные, негативные и нейтральные, учитывая, что у текста может быть только одна тональность;

Задача 2: определите авторство текста, выбрав между двумя авторами (они никогда не работали вместе);

Задача 3: определите тему текста, учитывая, что текст может относиться к нескольким темам одновременно.

Отметьте только один овал в каждом ряду.

	Мультилейбл	Мультикласс	Бинарная классификация
Задача 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Задача 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Задача 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Какой тип наивного байесовского классификатора вы бы выбрали для классификации текстов? * 1 балл

Отметьте только один овал.

- ☐ GaussianNB
- ☐ MultinomialNB
- ☐ BernoulliNB
- ☐ CategoricalNB

4. Что можно сказать об этом классификаторе? * 1 балл

	precision	recall	f1-score	support
pos	0.59	0.40	0.48	50
neg	0.55	0.72	0.62	50
accuracy			0.56	100
macro avg	0.57	0.56	0.55	100
weighted avg	0.57	0.56	0.55	100

Отметьте все подходящие варианты.

- ☐ Он хорошо находит примеры негативного класса среди всех, но хуже определяет, что пример именно негативный
- ☐ Он хорошо находит примеры позитивного класса среди всех, но хуже определяет, что пример именно позитивный
- ☐ В целом, негативный класс определяется хуже позитивного
- ☐ В целом, позитивный класс определяется хуже негативного
5. Как вы определите, какие признаки имеют наибольшее значение для логистической регрессии при принятии решений об отнесении к какому-либо классу? * 1 балл

Отметьте только один овал.

- ☐ Вызову функцию predict_proba() на тестовых данных
- ☐ Посмотрю на коэффициенты (атрибут coef_)
- ☐ Посмотрю на confidence values

6. Вы читаете статью, в которой написано: "Мы использовали метод k * 1 балл ближайших соседей (kNN) для определения тематики текстов". Что вы думаете о корпусе, который использовался для статьи? Выберите наиболее вероятный вариант

Отметьте только один овал.

- ☐ У авторов был большой корпус текстов, размеченный по темам
- ☐ Скорее всего, у авторов был неаннотированный корпус текстов
- ☐ Авторы разместили небольшой кусок неаннотированных данных

7. AUC вашего бинарного классификатора равна 0.46. Что это * 1 балл означает?

Отметьте только один овал.

- ☐ Что классификатор угадывает половину отрицательных примеров и половину положительных
- ☐ Что классификатор чаще путает позитивные примеры с негативными, чем предсказывает правильно
- ☐ Что значения TPR в среднем намного превосходят значения FPR

8. Соотнесите цель и название метода:

* 3 балла

Цель 1: уравновесить баланс классов в несбалансированных данных;

Цель 2: сделать так, чтобы отношение классов сохранилось во всех выборках (тренировочной, валидационной, тестовой);

Цель 3: избавиться от случаев деления на ноль при подсчете вероятностей.

Отметьте только один овал в каждом ряду.

	Стратификация	Сглаживание Лапласа	Присвоение весов
Цель 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Цель 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Цель 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Какая строчка на картинке **лишняя**? *

1 балл

```
1 from sklearn.feature_extraction.text import CountVectorizer
2
3 X_train = [...]
4 labels = [0, 1, 0, 1...]
5
6 vectorizer = CountVectorizer()
7 vectorizer.fit(X_train)
8 X_train_transformed = vectorizer.fit_transform(X_train)
9 X_train_transformed = vectorizer.transform(X_train)
```

Отметьте только один овал.

- ☐ 7
- ☐ 8
- ☐ 9
- ☐ Ни одна

10. Какой строчки на картинке **не хватает?** *

1 балл

```
1 from sklearn.naive_bayes import BernoulliNB
2 from sklearn.metrics import classification_report
3
4 # внимание: X - ЧИСЛОВЫЕ признаки, их не надо векторизовать
5 X_train = [...]
6 y_train = [0, 1, 0, 1...]
7 X_test = [...]
8 y_test = [0, 1, 0, 1...]
9
10
11 clf = BernoulliNB()
12 clf.fit(X_train, y_train)
13 print(classification_report(y_test, y_pred))
```

Отметьте только один овал.

- ☐ y = clf.predict(X_test)
- ☐ y_pred = clf.predict_proba(X_test)
- ☐ clf.predict(X_test)
- ☐ y_pred = clf.predict(X_test)

Компания Google не имеет никакого отношения к этому контенту.

Google Формы

