

ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Фонды оценочных средств по дисциплине «Методы машинного обучения»

Направление подготовки 01.04.02 Прикладная математика и информатика

## **1. Формируемые дисциплиной компетенции**

**ПК.8** Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач

### **Индикатор**

**ПК-8.1.** Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области

**ПК-8.2.** Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области

**ПК-8.3.** Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий

2. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания (в отсутствие утвержденных соответствующих локальных нормативных актов на факультете)

Вопросы к контрольным работам

1. Показатель  $X$  в классах  $K1$  и  $K2$  распределён нормально с параметрами: в  $K1$  математическое ожидание 2, стандартное отклонение 4; в  $K2$  математическое ожидание 3, стандартное отклонение 1. Выделить на числовой оси значений показателя  $X$  области отнесения байесовским классификатором к классам  $K1$  и  $K2$ . Априорные вероятности классов  $K1$  и  $K2$  равны 0.6 и 0.4 соответственно.
2. Каждый год варан подрастает на  $A\%$  от своего веса в начале года.  $A$  – случайная величина с известными матожиданием 5 и дисперсией 1 (одна и та же для всех варанов во все годы). В начале жизни каждый варан имеет вес 1. Построить байесовский классификатор для определения возраста варана (полных лет) по его весу, минимизирующий частоту ошибки. Предположить, что распознаваться будут «достаточно» взрослые вараны.
3. Выборка объектов из класса 1 и класса 2 определяется таблицами ниже. Указать тупиковые тесты.

	X1	X2	X3	X4					X1	X2	X3	X4	
Об.1	0	1	1	0					Об.1	0	1	0	0
Об.2	0	0	1	1					Об.2	1	0	1	0
Об.3	1	0	0	1					Об.3	0	1	0	0
Об.4	1	0	1	1					Об.4	1	1	0	0

4. Тестирование в банке системы распознавания для определения недобросовестных заёмщиков выявило связь между чувствительностью и ложной тревогой, показанную в таблице. Определить, приведёт ли эксплуатация системы к увеличению доходов банка. Определить возможный прирост дохода в расчёте на одну поданную заявку. Известно, что доход банка на одного заёмщика составляет 40000 денежных единиц, потери в результате отказа заёмщика от платежей составляют 120000 единиц. Доля недобросовестных заёмщиков составляет 7%.

Чувст.	Лож. Тр.
0.02	0.0001
0.12	0.003
0.23	0.05
0.38	0.12
0.47	0.16
0.58	0.19
0.67	0.23
0.78	0.34
0.89	0.52
0.97	0.72
1	0.87

5. В таблице даны значения переменных  $X$  и  $Y$  для четырёх экспериментов. Найти коэффициент корреляции и значения коэффициентов  $a$  и  $b$  для оптимальной по методу наименьших квадратов линейной модели  $Y=a + b \cdot X$ .

X	Y
0.12	52
0.23	37
0.35	17
0.46	2

6. Рассматривается задача классификации на два класса: положительный и отрицательный. В ходе тестирования классификатора получены следующие

результаты: полнота составляет 75%, общая точность составляет 80%. Какие значения может принимать точность?

7. Магазин собрал сведения о покупках (транзакции в файле). Были построены ассоциативные правила. Какое правило, содержащее в условии 2 элемента, имеет наибольшую поддержку?
8. Государственная избирательная комиссия зафиксировала результаты выборов по партиям и по регионам (таблица в файле). Требуется кластеризовать регионы по правилу  $k$ -средних для числа кластеров  $K$  от 1 до 12. Для каждого числа кластеров  $K$  найти максимальный радиус кластера. Построить график этой величины от  $K$ . На основании графика предположить, сколько групп регионов разумно выделить по итогам выборов.
9. В алгоритме вычисления оценок написать формулу для числа голосов, если система опорных множеств состоит из всех непустых подмножеств, а функция близости определяется только порогами  $\epsilon_1, \dots, \epsilon_n$ .
10. Обоснуйте способ построения всех тупиковых тестов через приведение системы тестовых уравнений к неупрощаемой ДНФ.