

Вариант 1

Правила

- Коллоквиум длится 1 час
- Можно пользоваться материалами лекций и своими записями
- Нельзя пользоваться прочими материалами в интернете
- Нельзя обсуждать задания с сокурсниками
- Компьютер можно использовать для выполнения базовых вычислений (арифметические операции, сортировка, поиск минимальных и максимальных элементов)

Задача 1

1. В таблице даны попарные расстояния между объектами из обучающей выборки. Проведите иерархическую кластеризацию с использованием complete-linkage расстояния между кластерами.

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0.0	1.5	5.0	4.0	2.5	0.5
x_2		0.0	4.0	0.5	3.5	2.0
x_3			0.0	6.0	2.0	1.0
x_4				0.0	5.5	4.5
x_5					0.0	1.0
x_6						0.0

Задача 2

На оси x заданы точки $[-5, -4, -3, 3, 4, 5]$. Также известны начальные параметры EM -алгоритма для разделения смеси нормальных распределений $K = 2, \mu_0 = 0, \sigma_0 = 1, \pi_0 = 0.5, \mu_1 = 0, \sigma_1 = 1, \pi_1 = 0.5$. Рассчитайте первую итерацию EM -алгоритма.

1. Необходимо вычислить начальное значение функции правдоподобия (можно воспользоваться `scipy.stats.norm` для расчетов), а так же новые значения γ, μ, σ, π . Нужно выписать все формулы и промежуточные расчеты.
2. Что произошло после первой итерации? Что нужно изменить в условии задачи при неизменных x , чтобы этого не случилось?

Задача 3

1. Напишите формулу критерия Silhouette. Дайте определение каждой составляющей формулы.
2. Приведите пример разбиения объектов в \mathbb{R}^2 на кластеры, при котором у некоторых объектов будет отрицательный силуэт.

Задача 4

1. Запишите функцию потерь в прямой задаче оптимизации для метода опорных векторов с квадратичными штрафами ξ_i^2 (используя отступ на i -м объекте $M_i = y^{(i)}g(x^{(i)})$). Какая регуляризация содержится в этом функционале?
2. Пусть N – количество объектов в обучающей выборке, S – количество опорных векторов, D – количество признаков, K – время расчета значения ядра для пары объектов. Оцените время выполнения предсказания
 - (a) Линейного SVM
 - (b) Ядерного SVM
3. Используя правила составления ядер покажите, что $k(u, v) = \langle u, v \rangle^d + c$ - ядро

Задача 5

1. Что такое Latent Factor Model для рекомендательных систем? Выпишите функционал качества и поясните смысл каждого элемента.
2. Предположим была обучена упрощенная Latent Factor Model: без корректировок на \bar{R}_i , \bar{R}_u и регуляризации. Для некоторых четырех фильмов было получено их скрытое представление:
 - $q_1 = (10, -1)$
 - $q_2 = (-2, 3)$
 - $q_3 = (3, 3)$
 - $q_4 = (-3, 6)$

Новый пользователь поставил фильмам 1, 2, 3 оценки 5, 1, 3 соответственно.

- (a) Найдите скрытый вектор p_u для этого пользователя (Hint: Normal Equation)
- (b) Посчитайте предсказание оценки этого пользователя для четвертого фильма \hat{R}_{u4}