

## Семинар 15

### Класификация на ИЧ и мас-спектри с метода на центроидите и линейната обучаваща машина

В този семинар се прилагат методите на центроидите и ЛОМ към реални спектрални данни. Работи се с три извадки от данни: (1) мас-спектри на алкани/алкени, (2) мас-спектри на алкани/циклоалкани и (3) ИЧ спектри на небензени/бензени.

**15.1. Класификация с метода на центроидите.** Прегорете материала от лекции 3 и 4. Припомнете си как се изчисляват различните разстояния между образите и центроидите, а също и как се взима решение един образ от кой клас е.

**Задача C15.1.** Стартирайте програмата `irmassclass.exe`. Отворете извадката от спектрални образи `alkaenes.set`. Разгледайте няколко образа от тази извадка, като навигирате със скролбара и двата радиобутона. Отворете този файл в текстовия редактор на windows и го разгледайте - отбележете си за спектрите на какви класове съединения става въпрос. Колко образа общо има в извадката? Колко от тях са от клас 0 и колко от клас 1? Колко образа има в обучаващата извадка и колко в тестващата извадка? Колко образа от клас 0 и клас 1 има в обучаващата извадка? А в тестващата извадка? Колко е размерността на образите?

**Задача C15.2.** С вече отворената извадка от спектрални образи `alkaenes.set` класифицирайте образ 13 от тестващата извадка като използвате различни степени в разстоянието на Минковски. От кой клас е образ 13? Класифицирайте образ 141 от тестващата извадка. От кой клас е образ 141?

**Задача C15.3.** В извадката от спектрални образи `alkaenes.set` класифицирайте образи 47 и 123 от тестващата извадка като използвате

разстояние на Хеминг и на Танимото при праг (*threshold*) 30% от максималния пик в спектъра. От кой клас са тези образи? С бутона `Report` получите информация за двоичните образи, в които се преобразуват образите 47 и 123. Използвайки тази информация, изчислете на ръка разстоянията до двата центроида от образи 47 и 123.

**Задача C15.4.** С помощта на бутона `Show Mean Patterns` получите двата центроида (средни вектора).

**Задача C15.5.** Отворете извадката от спектрални образи `alkcyes.set`. Със спектрите на тази извадка повторете задачи **C15.1** до **C15.4**.

**Задача C15.6.** Отворете извадката от спектрални образи `ir_benz.set`. Със спектрите на тази извадка повторете задачи **C15.1** до **C15.4**. При класифициране с разстоянията на Хеминг и Танимото как се изчисляват двоичните признаци?

**Задача C15.7.** За двата масспектъра по-долу кодирайте двоично двата спектъра с праг *Abundance* = 10% в интервала  $m/e = (41 - 44)$ . Изчислете разстоянията по Хеминг и Танимото между тях.



**15.2. Класификация с метода на ЛОМ.** Прегорете материала от лекция 6. Припомнете си как се изчисляват скаларното произведение на образи, тяхната големина и корекцията на тегловния вектор.

**Задача C15.8.** Отворете извадката от спектрални образи `alkscys.set`. Със спектрите на тази извадка проведете обучение на ЛОМ с начални коефициенти, равни на 1. Проверете