Семененко Марина Геннадьевна

**Модели и методы машинного обучения**

**Задача обучения по прецедентам**

Идея обучающихся машин принадлежит А. Тьюрингу.

Машинное обучение – процесс, в результате которого машина (компьютер) способна показывать поведение, которое в неё не было явно заложено.

Говорят, что компьютерная программа обучается на основе опыта E по отношению к некоторому классу задач T и меры качества P, если качество решения задач из T, измеренное на основе P, улучшается с приобретением опыта E.

Тест Тьюринга предложен Алланом Тьюрингом в 1950 году для определения способности машины проявлять интеллект, сравнимый с машиной.

**Основные понятия**

**Объекты и признаки**

fj: x >> Dj, j = 1, …, n – признаки объектов (features).

Типы признаков:

* Бинарный признак (0 или 1)
* Номинальный признак |Dj| < бесконечность (множество - конечное)
* Упорядочено-порядковый признак (множество – конечное и упорядоченное)
* Количественный признак Dj = R (вся числовая ось)

Вектор (f1(x), …, f2(x)) – признаковое описание объекта.

Матрица «объекты признаков» (features data) – в каждой строке конкретный объект, каждый столбце – конкретный признак по объекту строки.

x = (x1, x2, xd) принадлежит X = Q1 x Q2 x … x Qd

xj – j-й признак (свойство атрибут, предикативная переменная, feature) объекта r.

Если Qj конечно, то j-й признак – номинальный (категориальный или фактор).

Например, Qj = {Alfa Romeo, Audi, BMW, …, Lada}.

Если |Qj| = 2, то признак бинарный и можно считать, например, Qj = {0, 1}.

Если Qj конечно и упорядочено, то признак порядковый.

Например, Qj = {Beginner, Elementary, Intermediate, Advanced, Profeciency}.

Если Qj = R, то признак количественный.

Выход:

* Задача восстановления регрессии (аппроксимация)
* Задача классификация

**Задача обучения по прецедентам**

X – множество объектов

Y – множество ответов

y: X >> Y – неизвестная зависимость (target function)

Дано:

{x1, x2, x3, …, xi} – обучающая выборка (training sample)

y1, …, y(xi), I = 1, …, I – известные ответы (по обучающей выборке)

Найти:

a: X >> Y – алгоритм, решающую функцию (decision function), приближающую y на всём множестве X.

Весь курс ML – это конкретизация.

**Задача по классификации**

* Y {-1, +1} – классификация на 2 класса
* Y {1, …, M} – на M непересекающихся классов
* Y {0, 1}^M – на M классов, которых могут пересекаться

**Задача восстановления регрессии (regression)**

Y = R или Y = R^m

**Задача ранжирования**

Получить Y – конечное упорядоченное множество.