

ГЛОССАРИЙ

Агломеративные методы (методы «снизу вверх») — методы иерархической кластеризации, в которых построение иерархии происходит путем объединения мелких кластеров в более крупные.

Бинарный признак — признак, который может принимать два значения. Например, пол больного.

Выход — в задаче индуктивного обучения свойство объекта, которое необходимо научиться предсказывать по другим признакам (свойствам) объекта.

Дедуктивное (аналитическое) обучение — вид машинного обучения, в котором машина (компьютерная программа) выводит новые правила и факты на основе знаний, сформулированных экспертом.

Диаграмма рассеяния — графический способ представления данных об объектах, в котором два или три качественных свойства объектов рассматриваются как координаты x , y (x , y , z соответственно) и отображаются на плоскости (в пространстве соответственно).

Дивизивные методы (методы «сверху вниз») — методы иерархической кластеризации, в которых построение иерархии происходит путем разбиения кластеров на подкластеры.

Задача классификации (задача распознавания образов) — задача обучения с учителем, в которой выход есть качественный признак. Множество всех объектов с заданным выходом называется классом. Таким образом, множество всех объектов разбивается на непересекающиеся классы и требуется построить решающее правило (*классификатор*), которое по описанию объекта определяет, к какому классу он принадлежит.

Задача восстановления регрессии — задача обучения с учителем, в которой выход есть количественный признак. Требуется построить решающее правило (*регрессию*), которое по свойству объекта определяет выход.

Задача кластеризации — одна из задач обучения без учителя. Заключается в выявлении кластеров в заданном множестве объектов.

Иерархическая кластеризация (таксономия) — представление множества объектов в виде иерархии, на каждом уровне которой кластеры разбиваются на два или более подкластера. В вершине иерархии располагается кластер, содержащий все объекты. На нижнем ярусе располагаются кластеры, содержащие по одному объекту.

Индуктивное обучение — вид машинного обучения, в котором машина (компьютерная программа) строит общее правило по эмпирическим данным.

Качественный (номинальный) признак — признак, который может принимать конечное множество значений. Например, название заболевания.

Классификатор — решающее правило в задаче классификации.

Кластер — группа близких по значениям своих признаков объектов.

Количественный признак — признак, принимающий числовые значения. Например, температура тела.

Контрольная (тестовая) выборка — набор объектов (их описаний), с приписанным каждому значению выхода. По контрольной выборке оценивается обобщающая способность построенного ранее решающего правила.

Линейная регрессия — регрессия, линейно зависящая от параметров.

Машина опорных векторов — метод построения решающего правила, основанный на переходе от пространства исходных признаков в другое (спрямляющее) пространство, в котором решающее правило линейно, и поиске оптимальной разделяющей гиперплоскости.

Машинное обучение — процесс, в котором машина (компьютер) показывает поведение, которое в нее не было явно заложено (запрограммировано). Машинным обучением называют также науку, изучающую процессы обучения в искусственных системах.

Метод наименьших квадратов — метод восстановления регрессии, заключающийся в том, что в качестве регрессионных коэффициентов выбираются значения, достигающие минимум остаточной суммы квадратов. В случае линейной регрессии регрессионные коэффициенты являются решением системы нормальных уравнений.

Метод медиан — метод, решающий задачу кластеризации, аналогичный методу центров тяжести, где вместо последних рассматриваются медианы (объекты, наиболее близкие к центру тяжести).

Метод центров тяжести — итерационный метод, решающий задачу кластеризации. На каждой итерации вычисляются центры тяжести для каждого предварительно сформированного кластера и объекты перегруппировываются в соответствии с тем, к центру тяжести какого кластера данный объект расположен ближе всего.

Множественная регрессия — регрессия, зависящая более чем от одной входной переменной.

Нейронная сеть (многослойный перцептрон) — метод решения задач индуктивного обучения, основанный на принципе взаимодействия большого числа элементов (нейронов).

Обобщающая способность (качество) решающего правила — способность решающего правила правильно предсказывать выход для новых объектов, не вошедших в обучающую выборку.

Обучающая выборка — набор объектов (их описаний), с приписанным каждому значением выхода. По обучающей выборке в задаче обучения с учителем строится решающее правило.

Обучение без учителя — установление структурных свойств на множестве объектов, для которых известны их описания.

Обучение с учителем — построение решающего правила по обучающей выборке.

Объект (вход) — в задаче индуктивного обучения элемент эмпирических данных.

Описание объекта — набор значений признаков объекта (кроме выхода).

Оптимальная разделяющая гиперплоскость — в методе опорных векторов гиперплоскость, разделяющая два класса объектов, для которой расстояние от нее до ближайшего объекта из первого класса равно расстоянию до

ближайшего объекта из второго класса. При этом гиперплоскость должна занимать такое расположение, что эти расстояния максимальны.

Остаточная сумма квадратов — сумма квадратов отклонений для группы объектов, например, для всех объектов обучающей выборки.

Перцептрон — разновидность нейронной сети, разработанная Ф. Розенблаттом, а также устройство, реализующее эту сеть.

Признак (свойство) — в задаче индуктивного обучения некоторая характеристика объекта. Например, для объекта «пациент» свойством может являться пол, возраст, температура тела, название заболевания и т. д. У каждого объекта признакам приписаны некоторые значения.

Решающее правило (решающая функция) — закон, который по описанию объекта определяет значение выхода. Задача обучения с учителем заключается в построении решающего правила по обучающей выборке.

Регрессия — решающее правило в задаче восстановления регрессии.

Регрессионные коэффициенты — параметры решающей функции (регрессии) в задаче восстановления регрессии.

Система нормальных уравнений — система линейных уравнений для определения регрессионных коэффициентов линейной регрессии.