АНАЛИЗ ДИСКРИМИНАНТНЫЙ

Автор: словарь
31.08.2007 22:21 -

АНАЛИЗ ДИСКРИМИНАНТНЫЙ - группа методов многомерной статистики, предназначенных для 1) описания различий между классами и 2) классификации объектов, не входивших в первоначальную выборку обучающую (см.). Переменные (см.), используемые в А.Д., называются дискриминантными.

Для решения первой задачи строится пространство канонических дискриминантных функций, которые позволяют с максимальной эффективностью "разделить" классы. Для того чтобы разделить k классов, требуется не более (k - 1) канонических дискриминантных функций (например, для разделения двух классов достаточно одной функции, для разделения трех классов - двух функций и т.д.). Канонические дискриминантные функции можно рассматривать как аналог регрессии, построенной для целей классификации; дискриминантные (исходные) переменные являются в них переменными независимыми (см.). Для измерения абсолютного и относительного вклада дискриминантных переменных в разделение классов используются нестандартизированные и стандартизированные коэффициенты канонических функций.

В пространстве канонических дискриминантных функций можно также решать задачу классификации объектов, не принадлежавших к первоначальной выборке.

Для этого вычисляются расстояния от каждого "нового" объекта до геометрического "центра" каждого класса. Могут учитываться априорные вероятности принадлежности к классам, а также цена ошибок классификации.

Альтернативным подходом к классификации объектов является вычисление классифицирующих функций Фишера - по одной для каждого класса. Эти функции также можно рассматривать как аналог регрессионных уравнений с "независимыми" дискриминантными переменными. Объект относится к тому классу, для которого вычисленное значение классифицирующей функции является максимальным.

Первоначальная выборка, для каждого объекта из которой априорно известна классовая принадлежность и на основе которой строятся как классифицирующие, так и канонические дискриминантные функции, называется обучающей. По ней же оценивается эффективность решающих процедур. Для этого строится специальная классификационная матрица, которая показывает, к какому классу объект принадлежал априорно и в какой класс был классифицирован с помощью канонических дискриминантных или классифицирующих функций.

Метод А.Д. предъявляет к данным довольно строгие требования. В модели должно быть не менее двух классов, в каждом классе - не менее двух объектов из обучающей выборки, число дискриминантных переменных не должно превосходить объем обучающей выборки за вычетом двух объектов.

Дискриминантные переменные должны быть количественными и линейно независимыми (не должны коррелировать друг с другом). Выполнение этих требований проконтролировать достаточно легко. Для каждого класса требуется также приблизительное равенство матриц ковариации и многомерная нормальность распределения. Нарушение последнего требования может привести к ошибкам классификации в "пограничных" зонах, где вероятности принадлежности объекта к двум или нескольким классам приблизительно равны.

О.В. Терещенко