

 Самоорганизующиеся сети

Описание

Самоорганизующиеся сети позволяют выявлять кластеры (группы) входных векторов, обладающих некоторыми общими свойствами. Таким образом, они хорошо подходят для кластерного анализа.

К самоорганизующимся сетям относятся, в частности, сети Кохонена. Сети Кохонена представляют собой разновидность самоорганизующихся карт признаков, которые, в свою очередь, являются специальным типом нейронных сетей. Сеть Кохонена состоит из узлов, которые объединяются в кластеры. Наиболее близкие узлы соответствуют похожим объектам, а удаленные друг от друга — непохожим.

Также одним из типов самоорганизующихся сетей является нейронный газ. Нейронный газ — это алгоритм, позволяющий осуществлять адаптивную кластеризацию входных данных, то есть не только разделить пространство на кластеры, но и определить необходимое их количество, исходя из особенностей самих данных.

Для получения результирующих наборов требуется предварительное обучение узла.

Порты

Вход

-  Входной источник данных (таблица данных) — обязательный порт.

Требования к принимаемым данным

Поле будет запрещено к использованию, если:

- оно является дискретным и содержит всего одно уникальное значение;
- оно непрерывное и с нулевой дисперсией;
- оно содержит пропущенные значения.

Выход

-  Разбиение на кластеры. Таблица, состоящая из полей:

- **Номер кластера** — каждому объекту присвоен номер того кластера, в который он входит. Кластер состоит из узлов сети.
- **Номер узла** — каждому объекту присвоен номер узла сети, к которому он относится. Узлы в свою очередь объединяются в кластеры (поле *Номер кластера*).
- **Расстояние до узла** — евклидово расстояние до ближайшего узла, т.е. того, которому принадлежит объект.
- Все поля исходного набора данных, независимо от назначения поля.
-  Координаты узлов. Таблица, состоящая из полей:
 - **Номер узла** — перечислены номера сформированных узлов.
 - **Номер кластера** — перечислены номера сформированных кластеров.
 - Поля исходного набора данных, для которых значение параметра *Назначение* выбрано *Используемое*. В ячейках полей указаны веса узлов сети.

Мастер настройки

Мастер настройки включает в себя следующие группы параметров:

- Настройка входных столбцов. На данном этапе производится выбор полей для кластеризации:
 - Для полей, участвующих в кластеризации, выставить назначение *Используемое*.
 - Для прочих полей оставить *Не задано*.
- Настройки нормализации.
- Самоорганизующаяся нейронная сеть.
 - Тип самоорганизующейся сети:
 - Сеть Кохонена.
 - Нейронный газ. При выборе данной опции блок *Параметры сети* становится неактивным.
 - Параметры сети:
 - Топология сети: 1D, 2D прямоугольная, 2D гексагональная, 3D.
 - Размер сети по X. Параметр доступен для редактирования при любой топологии сети. Устанавливается значение целого типа в интервале от 2 до 100. По умолчанию — 3.
 - Размер сети по Y. Параметр доступен для редактирования при следующих топологиях сети: 2D прямоугольная, 2D гексагональная, 3D. Устанавливается значение целого типа в интервале от 2 до 100. По умолчанию — 3.
 - Размер сети по Z. Параметр доступен для редактирования при топологии сети 3D. Устанавливается значение целого типа в интервале от 2 до 100. По умолчанию — 3.
- **Random seed** — начальное число (целое, положительное), которое используется для инициализации генератора псевдослучайных чисел. Последовательность чисел генератора полностью определяется начальным числом. Если генератор повторно инициализируется с тем же начальным числом, он выдаст ту же последовательность чисел.

Параметр влияет на воспроизводимость результата обучения. Можно повторить результат обучения узла, если подать те же данные и выставить тот же random seed.

Для параметра доступны следующие команды:

- Всегда случайно — начальное число всегда будет случайным.
- Генерировать — сгенерируется новое начальное число.
- Копировать — в буфер обмена будет скопировано указанное значение.