УДК 681.3

О РЕШЕНИИ ЗАДАЧ КЛАССИФИКАЦИИ В МЕТОДАХ МАШИННОГО ОБУЧЕНИЯ

© 2018 Н. Е. Кравцова, А. П. Преображенский

Воронежский институт высоких технологий (г. Воронеж, Россия)

В статье проведен анализ некоторых подходов, полезных для решения задач классификации при использовании методов машинного обучения. Продемонстрированы возможности снижения размерности, ускорения работы алгоритмов.

Ключевые слова: машинное обучение, классификация, градиентный спуск.

Методы, связанные с машинным обучением, объединяют в себе разделы из большого числа дисциплин: теории оптимизации, численных методов, методов статистической обработки данных, дискретный анализ. Более новые подходы, например, data mining, базируются на них.[1]

Большое число задач сводятся к проблеме классификации. Это может быть распознавание образов, разграничение прав доступа между пользователями в компьютерной системе,

Одним из возможных инструментов может быть линейный классификатор. В нем проводится формирование разделяющей поверхности, имеющей линейную форму.

Если требуется разделить два класса, то происходит деление по двум полупространствам пространства признаков.

Поверхность будет кусочно-линейной, если будет проводиться анализ по большему числу классов.

Существуют и другие подходы. Например, кусочно-линейную поверхность для пространства признаков можно увидеть в методе ближайшего соседа, тогда, когда в каждом классе будет выбран один объект, будет реализация линейного классификатора.

Нервная клетка нейроном может рассматриваться как основа для того, чтобы провести обоснование линейного классификатора. Примеры из нейрофизиологии переложили на математические аналоги.

При линейной классификации может быть эффективной модель МакКаллока-Питтса [5].

При ее построении было установлено, что при помощи нейронных сетей есть возможности достичь тех же результатов, что и для дискретных устройств с конечной памятью, при помощи нейронных сетей. То есть, можно реализовать всевозможные логические операции.

В такой модели не всегда можно достичь требуемой гибкости в процессе обучения, когда решаются определенные задачи.

Это связано с тем, что переход из одного состояния в другое связан с порогом. И если порог не преодолен, то можно не получить сигнал на выходе. Существует определенная инертность или рефрактерность.

При помощи градиентных методов можно провести настройку многослойных сетей. В методе обратного распространения ошибок эффективным будет вычисление градиента, поскольку проводится аналитическое дифференцирование суперпозиции. При этом можно столкнуться с тем, что будет медленная сходимость.

Для линейного классификатора могут быть использованы градиентные методы.

Градиентный спуск можно реализовать двумя способами:

1. Пакетным, при этом просмотр обучающей выборки осуществляется целиком для каждой из итераций,

2. Стохастическим, при котором делается случайный выбор лишь одного объекта для каждой из итераций алгоритма. Существуют возможности улучшения алгоритма за счет эвристик, когда задается определенный закон для вероятностей.

Градиентный спуск используется при обучении персептрона, Правила Хебба позволяют неверном выходном сигнале увеличивать или уменьшать веса на входах, равных 1 (при выходном сигнале 0 или 1, соответственно). Дальнейшим их развитием яв-

[1] Кравцова Нина Евгеньевна – Воронежский институт высоких технологий, студент, kravtsovanina@yandex.ru.
Преображенский Андрей Петрович – Воронежский институт высоких технологий, д. т. н., профессор, app@vivt.ru.

ляется дельта-правило и обобщенное дельта-правило.

Для градиентных методов в определенных случаях характерна медленная сходимость, в этой связи полезными с точки зрения практики могут быть использованы или метод Левенберга-Марквардта или адаптационные процедуры в методе сопряженных градиентов.

Метод k-средних также может быть использован для того, чтобы осуществить кластеризацию.

В нем рассматривается суммарное квадратичное отклонение между точками в кластерах и центрами таких кластеров. Его требуется минимизировать.

Следует отметить, что для алгоритмов глубокого обучения такой подход может быть использован для того, чтобы осуществлять фильтрацию, тогда формируются ядра свертки.

Веса в методе k-средних могут быть удалены с использованием редукции [9] при условии, что средние значения значительно их превосходят.

Следует следить за погрешностью сети, чтобы она при этом не возрастала. Но при этом процесс обучения должен быть завершен, поскольку текущий вес может оказывать влияние при своем изменение на то, какая чувствительность сети.

Градиент не всегда демонстрирует степень чувствительности, поскольку вблизи минимума будет малое изменение градиента. В этой связи требуется рассматривать не первые, а вторые производные.

В теории распознавания образов задача, связанная с классификацией может достаточно наглядно быть продемонстрирована в рамках байесовского подхода [6-8].

В нем исходят из того, что существуют возможности представления алгоритма классификации явным образом, тогда, если известны плотности распределения по классам.

Эти плотности могут быть определены различным образом: непараметрическим, параметрическим и за счет того, что смеси распределений оцениваются исследователями.

Первый из подходов исходит из того, что плотности распределения по классам для окрестностей анализируемых объектов локальным образом оцениваются.

В случае равновероятных и равнозначных классов в уравнении разделяющей поверхности используется расстояние Махаланобиса и будет равенство ковариационных матриц. Классификатор ближайшего среднего может считаться оптимальным.

Евклидова метрика будет равна расстоянию Махаланобиса для случая одинаковых дисперсий признаков, если они независимы. Тогда оптимальное правило будет связано с тем классом, который ближайший к центру.

Для того, чтобы повысить устойчивость ковариационной матрицы можно применить эвристику Фишера, что не всегда позволяет полноценным образом решить задачу, вследствие близости к нулю собственных значений.

Другой подход исходит из того, что наименее значимые признаки будут отбрасываться. Один из способов - редукция размерности.

При этом можно уменьшить время анализа, поскольку обращаются матрицы, размерность которых не превосходит 2. Подобный "жадный" алгоритм эффективен для случаев с небольшим числом признаков.

Метод главных компонент является еще одним подходом, в котором выбираются наиболее информативные признаки, но их общее число является минимальным.

Плотность распределения выборки для параметрического подхода считается известной с точностью до параметра.

При этом принцип максимума правдоподобия [3, 4] позволяет сделать оценку вектора параметров. Для частного случая многомерных нормальных всех классов может быть применен нормальный дискриминантный анализ, позволяющий аналитическое решение задачи.

Не всегда прямое применение принципа максимума правдоподобия позволяет прийти к оптимизационной задаче в удобном для решения виде. Для того, чтобы упростить преобразования можно использовать алгоритм EM (expectation-maximization) [2]. В нем прибегают к тому, что используют вектор скрытых переменных.

Есть зависимость скорости сходимости к решению и его значений от того, какое начальное приближение [1].

Проведение выбора по начальному приближению и числу компонентов можно осуществлять за счет последовательного добавления компонент.

Выводы. При выборе соответствующего метода классификации необходимо исходить из его требуемой точности, быстродействия, характеристик входных данных.

ЛИТЕРАТУРА

1. Айвазян С. А. Прикладная статистика: классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика. – 1989.
2. Dempster A. P. Maximum likelihood from incomplete data via the EM algorithm / A. P. Dempster, N. M. Laird, D. B. Rubin // J. of the Royal Statistical Society, Series B. – 1977. – no. 34. – Pp. 1-38.
3. Закс Ш. Теория статистических выводов / Ш. Закс. – М.: Мир. – 1975.
4. Лагутин М. Б. Наглядная математическая статистика / М. Б. Лагутин. – М.: П-центр. – 2003.
5. McCulloch W. S. A logical calculus of ideas immanent in nervous activity / W. S.McCulloch, W.Pitts // Bulletin of Mathematical Biophysics. – 1943. – no. 5. – Pp. 115-133.
6. Parzen E. On the estimation of a probability density function and mode / E. Parzen // Annals of Mathematical Statistics. – 1962. – Vol. 33. – Pp. 1065-1076. (http://citeseer.ist.psu.edu/parzen62estimation.html).
7. Rosenblatt M. Remarks on some nonparametric estimates of a density function / M. Rosenblatt // Annals of Mathematical Statistics. – 1956. – Vol. 27. – no. 3. – Pp. 832-837.
8. Хардле В. Прикладная непараметрическая регрессия / В. Хардле. – М.: Мир. – 1993.
9. Шурыгин А. М. Прикладная стохастика: робастность, оценивание, прогноз / А. М. Шурыгин. – М.: Финансы и статистика. – 2000.

ABOUT THE DECISION OF PROBLEMS OF CLASSIFICATION METHODS IN MACHINE LEARNING

2018 N. E. Kravtsova, A. P. Preobrazhensky

Voronezh Institute of high technologies (Voronezh, Russia)

This paper analyzes some approaches that are useful for solving classification problems using machine learning methods. The possibilities of reducing the dimensionality and speeding up the algorithms are demonstrated.

Key words: machine learning, classification, gradient descent.