

Министерство образования Республики Беларусь

Учреждение образования
**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ**

Факультет Компьютерных сетей и систем

Кафедра Информатики

РЕФЕРАТ

по курсу «Интеллектуальный анализ информации»

Многоклассовый дискриминантный анализ

Студент:
гр. 758641
Ярош Г.И.

Проверил:
Ивашенко В.П.

Минск, 2018

СОДЕРЖАНИЕ

МНОГОКЛАССОВЫЙ ДИСКРИМИНАНТНЫЙ АНАЛИЗ	3
1. Введение	3
2. Линейный дискриминантный анализ для двух классов.....	3
3. Дискриминантный анализ нескольких классов	4
4. Недостатки и ограничения дискриминантного анализа	4
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	6

МНОГОКЛАССОВЫЙ ДИСКРИМИНАНТНЫЙ АНАЛИЗ

1. Введение

Метод линейного дискриминантного анализа используется для нахождения линейных комбинаций признаков, наилучшим образом разделяющих два или более класса объектов или событий. Полученная комбинация может быть использована в качестве линейного классификатора или для сокращения размерности пространства признаков перед последующей классификацией.

ЛДА тесно связан с дисперсионным анализом и регрессионным анализом, также пытающимися выразить какую-либо зависимую переменную через линейную комбинацию других признаков или измерений. В этих двух методах зависимая переменная — численная величина, а в ЛДА она является величиной номинальной (меткой класса). Помимо того, ЛДА имеет схожие черты с методом главных компонент и факторным анализом, которые ищут линейные комбинации величин, наилучшим образом описывающие данные.

Для использования ЛДА признаки должны быть непрерывными величинами, иначе следует использовать анализ соответствий.

Оригинальный дихотомический дискриминантный анализ разработал сэр Роналд Фишер в 1936. Он отличается от дисперсионного анализа или многофакторного дисперсионного анализа, которые используются для предсказания одной (дисперсионный анализ) или нескольких (многофакторный дисперсионный анализ) непрерывных зависимых переменных по одной или более независимой качественной переменной. Анализ дискриминантных функций полезен для определения, является ли множество переменных эффективным в предсказании принадлежности категории.

В общем случае линейный дискриминантный анализ применяется для разделения объектов на два класса. Однако существуют подходы, которые позволяют применить ЛДА для классификации по многим классам объектов.

2. Линейный дискриминантный анализ для двух классов

Для каждого образца объекта или события с известным классом Y рассматривается набор наблюдений x (называемых ещё признаками, переменными или измерениями). Набор таких образцов называется обучающей выборкой (или набором обучения, обучением). Задачи классификации состоит в том, чтобы построить хороший прогноз класса Y для всякого так же распределённого объекта (не обязательно содержащегося в обучающей выборке), имея только наблюдения x .

Рассмотрим случай с двумя классами $Y = (0, 1)$. В линейном дискриминантном анализе предполагается, что плотности распределения объектов обоих классов

соответствуют нормальному распределению. Также изначально известны средние значения объектов для двух классов, а также их матрицы ковариации. Строится следующая функция правдоподобия:

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x} - \vec{\mu}_0) + \ln |\Sigma_0| - (\vec{x} - \vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x} - \vec{\mu}_1) - \ln |\Sigma_1| > T$$

где μ_0, μ_1 - средние значения плотностей вероятностей для обоих классов, Σ_0, Σ_1 – матрицы ковариации для первого и второго классов, T - некоторое пороговое значение.

Суть данной функции правдоподобия является то, что объект x относится к первому классу тогда, когда выполняется неравенство. Иначе данный объект относится ко второму классу.

При равенстве матриц ковариации для обоих классов, данная функция правдоподобия сводится к сравнению скалярного произведения с пороговым значением:

$$\vec{w} \cdot \vec{x} > c$$

$$\vec{w} = \Sigma^{-1} (\vec{\mu}_1 - \vec{\mu}_0)$$

$$c = \frac{1}{2} (T - \vec{\mu}_0^T \Sigma_0^{-1} \vec{\mu}_0 + \vec{\mu}_1^T \Sigma_1^{-1} \vec{\mu}_1)$$

Данное следствие означает, что критерием вхождения некоторого объекта x в класс Y является функцией линейной комбинации от значений вектора x .

3. Дискриминантный анализ нескольких классов

При решении задачи классификации по нескольким классам с помощью линейного дискриминантного анализа могут быть применены следующие подходы:

- Для каждого класса можно положить, что все остальные точки принадлежат к другому классу. В таком случае мы получаем N задач бинарной классификации, где N – количество классов. Для решения данной задачи строятся N дискриминантных функций для каждого класса, которые комбинируются для получения конечного решения.
- Можно строить дискриминантные функции для каждой пары классов. В таком случае необходимо решить $N(N-1)$ задач классификации.

4. Недостатки и ограничения дискриминантного анализа

Предположения дискриминантного анализа являются теми же, что и для многофакторного дисперсионного анализа. Анализ весьма чувствителен к выбросам и размер наименьшей группы должен быть больше, чем число переменных в исходных объектах переменных.

- Нормальное распределение. Параметры исходных объектов должны быть нормально распределены.

- Однородность ковариации. Предполагается, что матрицы ковариации по всем входным переменным однородны.
- Корреляция между средним значением и дисперсией: Сила предсказания может уменьшаться с увеличением зависимости между средними значениями по совокупности объектов и дисперсиями.
- Независимость. Объекты предполагаются случайно распределёнными и оценка по одной переменной для объекта независима от оценки по другой переменной.

Практическое применение метода ограничивает тот факт, что значения средних и ковариаций классов в большинстве случаев неизвестны. Они могут быть оценены по обучающей выборке. Однако такие оценки среднего и ковариаций при подстановке в дискриминантную функцию не гарантируют ее оптимальность.

Также ограничением в применении описанного метода является то, что количество объектов в выборке должно превосходить количество переменных для объекта. В таком случае матрица ковариаций имеет не полный ранг и не может быть обращена. Для преодоления данной проблемы может использоваться псевдообратная матрица вместо обратной.

Во многих практических случаях линейный дискриминантный анализ не применим из-за невозможности линейно разделить изучаемые классы. В таком случае исходные объекты могут отображаться в пространство большей размерности и иметь другое количество переменных. Линейная классификация в таком пространстве эквивалентная нелинейной классификации в исходном пространстве.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Линейный дискриминантный анализ. [Электронный ресурс]. – Электронные данные. – Режим доступа: https://ru.wikipedia.org/wiki/Линейный_дискриминантный_анализ.

[2] Линейный дискриминантный анализ [Электронный ресурс]. – Электронные данные. – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Линейный_дискриминантный_анализ.

[3] Дискриминантный анализ [Электронный ресурс]. – Электронные данные. – Режим доступа: <http://statsoft.ru/home/textbook/modules/stdiscan.html#discriminant>.