

Метод LMNN обучения метрики

Лебедь Федор Сергеевич

МГУ имени М.В. Ломоносова, факультет ВМК, кафедра ММП

Почему метрика важна



- Метрика = модель пространства.
- Метрические алгоритмы работают с пространством, порожденным метрикой.

Как выбрать «хорошую» метрику

- Разработано огромное количество функций сходства для различных объектов
- Все функции сходства параметризованы
- Можно считать геодезическое расстояние
- В ручную уже никак, нужен алгоритм!

Классическое семейство метрик

$$d_A^2(x, y) = (x - y)^T A (x - y), \quad A \succeq 0$$

(Некоторые ошибочно называют расстоянием Махаланобиса)

Геометрическая интерпретация

По разложению Холецкого

$$d_A^2(x, y) = \|Ux - Uy\|^2, \quad A = U^T U$$

Воу-воу, это что, просто линейное преобразование пространства?
Да.

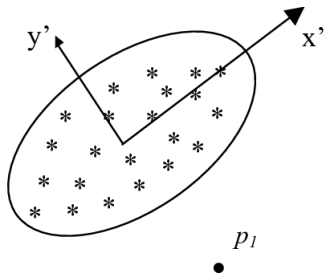


Рис. 1: Вот так выглядит окрестность в d_A^2

Слишком просто, давай сложнее

Чтобы решить проблему неоднородности пространства можно использовать мультиметрические методы.

- Припишем каждой точке пространства тензор
- В каждой точке пусть растягивается в соответствии с этим тензором
- Что-то слишком сложно

Можно попроще

- Каждому классу дадим свою метрику
- Расстояние до объекта = расстояние по метрики его класса
- One-vs-All сводит к LMNN, ура умеем решать!

LMNN заключается в выборе A для $d_A^2(x, y)$ так, чтобы

- Объекты одного класса были рядом
- Объекты разных классов были далеко
- Да ты что?!

Как решать?

- Для каждого объекта выбираем целевых соседей (до обучения)
- Ребят не из класса объекта называем самозванцами
- Боремся за то, чтобы выпихнуть самозванцев из окрестности целевых соседей
- Соберем задачу минимизации

Уезжайте...

$$\min_M \sum_{i,j \in N_i} d(\vec{x}_i, \vec{x}_j) + \sum_{i,j,l} \xi_{ijl}$$

$$d(\vec{x}_i, \vec{x}_j) + 1 \leq d(\vec{x}_i, \vec{x}_l) + \xi_{ijl} \quad \forall i,j \in N_i, l, y_l \neq y_i$$

$$\xi_{ijl} \geq 0$$

- Первая сумма минимизирует расстояние до целевых соседей
- Вторая сумма минимизирует нежесткость
- Неравенства выпихивают самозванцев из окрестности
- И конечно условие нежесткости

- Заменим ξ на hinge loss, дифференцируемость это убьет, но не сильно.
- И почему это эквивалентная замена? Вспоминай ММРО!

Параметров стало нормально, решаем итеративно, градиентным спуском.

Кукарек?

Да, в хорошем смысле

