

Практическая работа №9: «Метод опорных векторов».

Оглавление

Цель работы	1
Задачи работы	1
Перечень обеспечивающих средств.....	1
Общие теоретические сведения	1
Бинарная классификация.....	1
Линейная неразделимость	3
Регрессия	3
Задание	4
Требования к отчету	4
Литература	4

Цель работы

Получить практические навыки решения задач регрессии и классификации с помощью метода опорных векторов.

Задачи работы

1. Сравнить несколько моделей для решения задачи регрессии методом опорных векторов.
2. Сравнить несколько моделей для решения задачи классификации методом опорных векторов.

Перечень обеспечивающих средств

1. ПК.
2. Учебно-методическая литература.
3. Задания для самостоятельного выполнения.

Общие теоретические сведения

Бинарная классификация

Данные: элементы (X_i, y_i) , где $X_i = (x_{i1}, x_{i2}, \dots, x_{iK})$, $y_i \in \{Y_0, Y_1\}$,

$i = 1, \dots, N$ (размер набора данных).

Задача: Найти такую функцию $f(x)$, чтобы $f(x_i) \approx y_i$ для всех i .

Основная идея:

Построить гиперплоскость (фигуру размерность $K-1$), которая наиболее эффективно разделяет классы – разделяющую гиперплоскость:

$$\langle \vec{w}, \vec{X} \rangle - b = 0, \text{ где } \vec{w}, \vec{X} \in \mathbb{R}^K.$$

Опорный вектор – элемент класса, который находится ближе всего к разделяющей гиперплоскости.

Зазор (отступ) – сумма расстояний между опорными векторами и разделяющей гиперплоскостью.

Гиперплоскость $\langle \vec{w}, \vec{X} \rangle - b = 0$ разделяет K -мерное пространство на две части:

$$\begin{cases} \langle \vec{w}, \vec{X} \rangle - b < 0 \\ \langle \vec{w}, \vec{X} \rangle - b > 0 \end{cases}$$

Гиперплоскость выполняет роль классификатора для (X_i, y_i) , если

$$\begin{cases} \langle \vec{w}, \vec{X}_i \rangle - b < 0, \quad \forall y_i = Y_0 \\ \langle \vec{w}, \vec{X}_i \rangle - b > 0, \quad \forall y_i = Y_1 \end{cases}$$

$$\begin{cases} \langle \vec{w}, \vec{X}_i \rangle - b < 0, \quad \forall y_i = -1 \\ \langle \vec{w}, \vec{X}_i \rangle - b > 0, \quad \forall y_i = +1 \end{cases}$$

$$f(X) = \text{sign}(\langle \vec{w}, \vec{X} \rangle - b)$$

Мы хотим выбрать \vec{w} , чтобы максимизировать зазор.

Можно показать, что зазор равен $\frac{2}{\|\vec{w}\|}$.

$$\frac{2}{\|\vec{w}\|} \rightarrow \max$$

$$\|\vec{w}\| \rightarrow \min$$

$$\|\vec{w}\|^2 \rightarrow \min$$

$$\begin{cases} \langle \vec{w}, \vec{X}_i \rangle - b < 0, \quad \forall y_i = -1 \\ \langle \vec{w}, \vec{X}_i \rangle - b > 0, \quad \forall y_i = +1 \end{cases}$$

$$y_i(\langle \vec{w}, \vec{X}_i \rangle - b) > 0$$

$$\begin{cases} \|\vec{w}\|^2 \rightarrow \min \\ y_i(\langle \vec{w}, \vec{X}_i \rangle - b) \geq 0 \end{cases}$$

Такая задача оптимизации решается методом множителей Лагранжа.

Линейная неразделимость

Для «искривления» пространства элементов $X \in \mathbb{R}^K$ вводится отображение:

$X \rightarrow \varphi(X)$, где $\varphi(X) \in \mathbb{R}^M$, $M > K$.

Задача оптимизации принимает вид:

$$\begin{cases} \|\vec{w}\|^2 \rightarrow \min \\ y_i(\langle \vec{w}, \varphi(\vec{X}_i) \rangle - b) \geq 0 \end{cases}$$

Регрессия

Данные: элементы (X_i, y_i) , где $X_i = (x_{i1}, x_{i2}, \dots, x_{iK})$, $y_i \in \mathbb{R}$.

Задача:

Найти такие значения a и b , чтобы функция $f(x) = ax + b$ как можно точнее аппроксимировала y , т.е. чтобы $f(x_i) \approx y_i$ для всех i .

Задание

Пояснение

Для сохранения результатов данной работы вам понадобится файл `ipynb`. Если требуется, для удобства можно создать также второй файл формата `doc/docx`. Названия файла или файлов должны иметь вид «*Фамилия – задание 9*».

Часть 1

- Обновите свой репозиторий, созданный в практической работе №1, из оригинального репозитория:
https://github.com/mosalov/Notebook_For_AI_Main.

Часть 2

- Откройте свой репозиторий в Binder (<https://mybinder.org/>).
- Откройте файл «2021 Весенний семестр\task2.ipynb».
- Изучите, при необходимости – выполните повторно, приведённый в файле код.
- Выполните два задания, приведённых в ячейках в конце ноутбука.
- Сохраните код в `ipynb`-файле. При необходимости пояснения опишите в `doc/docx`-файле.

Требования к отчету

Готовые файлы загрузите в свой репозиторий, созданный в практическом задании №1 по пути: «Notebook_For_AI_Main/2021 Весенний семестр/Практическое задание 2/» и сделайте пул-реквест.

Литература

1. <https://habr.com/ru/post/105220/>
2. www.machinelearning.ru/wiki/index.php?title=Метод_опорных_векторов
3. <https://nuancesprog.ru/p/9573/>
4. <https://scikit-learn.org/stable/modules/svm.html>