VC Размерность и ее применение

к.ф.-м.н. А.Т. Вахитов

February 19, 2016

О чем мы поговорим

- Анализ конечной выборки: число дихотомий, функция роста, точка разрыва
- VC размерность: оценка отличия ошибки по выборке от полной ошибки
- PCA (метод главных компонент) и его приложения в машинном обучении
- Шум и отклонение при машинном обучении

На основе: Y. Abu-Mostafa, Learning From Data (Caltech online)

Было в прошлый раз

Задача обучения:
Необходимо найти наилучшее $h \in H$ из допустимого множества гипотез
Для гипотезы h определим

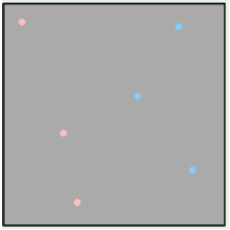
- in sample error (ошибка по выборке) $E_{in}(h) = \nu$
- out of sample error (ошибка по пространству) $E_{out}(h) = \mu$

Аналогия приводит нас к следующей форме нер-ва Бернштейна:

$$P(|E_{in}(h) - E_{out}(h)| > \epsilon) \leq 2e^{-2\epsilon^2 N}$$

Анализ конечной выборки: число дихотомий

Задача классификации. Пусть есть выборка $x_1, \dots, x_N$, а также набор бинарных значений $y_1, \dots, y_N$, задающих принадлежность элемента одному из двух классов.



Сколько вариантов $y_1, \dots, y_N$?

Анализ конечной выборки: число дихотомий (2)

Обозначим как $|H(x_1, \ldots, x_N)|$ число дихотомий (т.е. различных наборов $y_1, \ldots, y_N$, которое сможет породить наш набор гипотез H.

Тогда $|H(x_1, \ldots, x_N)| \leq 2^N$.

В то же время само множество гипотез H может быть бесконечно.

Пример: $X =$точки на плоскости, $H =$ приямые на плоскости

Выборка из трех точек x_1, x_2, x_3, $|H(x_1, x_2, x_3)| = 2^3 = 8$.

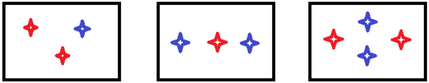
Функция роста $m_H(N)$

Функция роста $m_H(N)$ - для любых N точек $x_1, \ldots, x_N$

$$m_H(N) = max_{x_1,\ldots,x_N} |H(x_1, \ldots, x_N)|$$

Ясно, что $m_H(N) \leq 2^N$
Для перцептрона и $x_i \in \mathbb{R}^2$

Пример функции роста $m_H(N)$



$$m_H(3) = 8 = 2^3 \quad m_H(4) = 14 < 2^4$$

Не считаем множества меры 0

Примеры

- Гипотезы: лучи на множестве точек: прямая $m_H(N) = ?$
- Гипотезы: интервалы на множестве точек: прямая $m_H(N) = ?$
- Гипотезы: выпуклые множества на множестве точек: плоскость $m_H(N) = ?$

Точка разрыва

Если для какого-то $k \in \mathbb{N}$ $m_H(k) < 2^k$ то k называется *точкой разрыва*

Пример: $k = 4$ для перцептрона на плоскости

- Гипотезы: лучи на множестве точек: прямая; точка разрыва?
- Гипотезы: интервалы на множестве точек: прямая; точка разрыва?
- Гипотезы: выпуклые множества на множестве точек: плоскость; точка разрыва?

Задача

Известно, что точка разрыва 2. Сколько дихотомий может быть для трех элементов в выборке?

- $2^3 = 8$ дихотомий не реализуемы
- для каждой пары элементов $2^2 = 4$ реализуемы

Рассмотрим все возможные $m_H(3) = 8, 7, 6, 5, 4, 3, \ldots$?

Ответ задачи

Ответ задачи

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

Точка разрыва

Теорема: наличие (конечной) точки разрыва означает наличие полиномиального ограничения на функцию роста $m_H(N)$

Полиномиальное ограничение для функции роста

	# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\dots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
S_1	α	$+1$	$+1$	$\dots$	$+1$	$+1$
		-1	$+1$	$\dots$	$+1$	-1
		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	-1	-1
		-1	$+1$	$\dots$	-1	$+1$
S_2: S_2^+	β	$+1$	-1	$\dots$	$+1$	$+1$
		-1	-1	$\dots$	$+1$	$+1$
		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	$+1$	$+1$
		-1	-1	$\dots$	-1	$+1$
S_2: S_2^-	β	$+1$	-1	$\dots$	$+1$	-1
		-1	-1	$\dots$	$+1$	-1
		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		$+1$	-1	$\dots$	$+1$	-1
		-1	-1	$\dots$	-1	-1

Выпишем все возможные дихотомии. Разделим на 2 ситуации: для x_N есть 2 варианта (+1 и -1) либо 1 вариант (+1 либо -1)

Полиномиальное ограничение для функции роста

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\ldots$	$+1$	$+1$
			-1	$+1$	$\ldots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	-1	-1
			-1	$+1$	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	$+1$
			-1	-1	$\ldots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	$+1$
S_2			-1	-1	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	-1	-1

$B(N, k)$ максимальное число дихотомий для выборки размера N при наличии точки разрыва k

Докажем: $B(N, k) \leq f(N)$, $f = O(N^r)$

Индукция: для $N = k$ выполнено. Пусть верно для $N - 1 \geq k$, докажем для N.

Доказательство

Из таблицы: $B(N,k) = \alpha + 2\beta$

Докажем $\alpha + \beta \leq B(N-1,k)$

Очевидно (индукционная гипотеза), так как $\alpha + \beta$ - число дихотомий для $N-1$

Докажем $\beta \leq B(N-1,k-1)$

Очевидно, так как если есть поднабор размера k из $x_1, \ldots, x_{N-1}$, для которого есть все 2^{k-1} дихотомий, то для всех переменных будет 2^k дихотомий, что противоречит условию.

В итоге, доказано:

$$B(N,k) \leq B(N-1,k) + B(N-1,k-1).$$

Ограничение для $B(N, k)$

Теорема

$$B(N, k) \leq \sum_{i=0}^{k-1} C_N^i$$

Доказательство: индукция
индукционный шаг:

$$\sum_{i=0}^{k-1} C_N^i = \sum_{i=0}^{k-1} C_{N-1}^i + \sum_{i=0}^{k-2} C_N^i$$

Итого: для функции роста

Полиномиальное ограничение

$$m_H(N) \leq \sum_{i=0}^{k-1} C_N^i$$

Примеры оценок:

- Гипотезы: лучи на множестве точек: прямая; $m_H(N) = N + 1 \leq N + 1$
- Гипотезы: интервалы на множестве точек: прямая; $m_H(N) = \frac{N^2}{2} + \frac{N}{2} + 1 \leq \frac{N^2}{2} + \frac{N}{2} + 1$
- Гипотезы: перцептроны на множестве точек: плоскость; $m_H(N) = ? \leq \frac{N^3}{6} + \frac{5N}{6} + 1$

Итого: для функции роста

Покажем, что можно использовать $m_H(N)$ вместо M
Ключевые соображения: представить $E_{in} - E_{out}$ с помощью случайной выборки других N значений, т.е. выразить $E_{in} - E_{out}$ с помощью $E_{in} - E_{in2}$

Неравенство Вапника - Червоненкиса (VC)

$$P(|E_{in}(g) - E_{out}(g)| > \epsilon) \leq 4m_H(2N)e^{-\frac{1}{8}\epsilon^2 N}$$

Размерность VC

d_{VC} = наибольшее N, для которого $m_H(N) = 2^N$ ($d_{VC} = k - 1$ для k - точка разрыва)

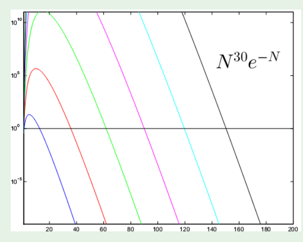
- Гипотезы: лучи на множестве точек: прямая; $d_{VC} = 1$
- Гипотезы: перцептроны на множестве точек: плоскость; $d_{VC} = 3$
- Гипотезы: выпуклые множества на множестве точек: плоскость; $d_{VC} = \infty$

Следствие неравенства Вапника - Червоненкиса (VC)

Неравенство VC:

$$P(|E_{in}(g) - E_{out}(g)| > \epsilon) \leq 4 m_H(2N) e^{-\frac{1}{8}\epsilon^2 N}$$

Зависимость N от d: рассмотрим $N^d e^{-N}$



Зависимость $N^d e^{-N}$ для $d = 5, 10, 15, 20, 25, 30$

Применения неравенства Вапника - Червоненкиса (VC)

Неравенство VC:

$$P(|E_{in}(g) - E_{out}(g)| > \epsilon) \leq \delta$$

$$\delta = 4m_H(2N)e^{-\frac{1}{8}\epsilon^2 N} \Rightarrow \epsilon = \sqrt{\frac{8}{N} ln \frac{4m_H(2N)}{\delta}}$$

С вероятностью $1 - \delta$, $|E_{out} - E_{in}| \leq \Omega(N, H, \delta)$

PCA (Principal Component Analysis) -Метод главных компонент

Есть набор векторов $x_1, \ldots, x_N \in \mathbb{R}^q$
Задача: построить базис из $s << q$ элементов $b_1, \ldots, b_s$, в котором можно представить x_i с минимальной ошибкой
Обозначим как $\hat{x}_i$ проекцию x_i на подпространство, заданное этим базисом

$$\sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \to \min$$

PCA: детали алгоритма

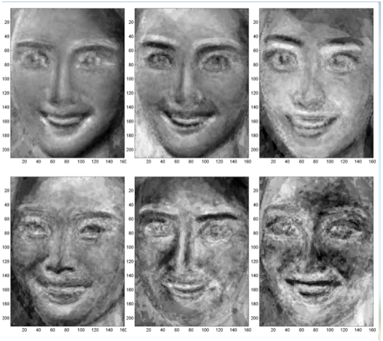
$x_1, \ldots, x_N$

1. центрировать ($\bar{x}_i = x_i - \bar{x}$, $\bar{x} = \frac{1}{N}\sum x_i$)
2. составить матрицу $X = (x_1\ x_2\ \ldots\ x_N)^T$
3. вычислить $X = USV'$ - $svd-$ разложение
4. результат - первые s столбцов матрицы V.

Пример: распознавание лиц: eigenfaces (Turk, Pentland 1991)

Класс = человек, x_i = изображение лица

Первые 6 "главных векторов" для набора лиц Miss Korea 2013



Интересный пост про PCA и лица корейских моделей