Сборник задач по курсу "Машинное обучение"

авторы: Кантонистова Е.О., Титов В.В., Широков А., Поликарпов К.

19 октября 2020 г.

Содержание

1 Матричное дифференцирование

1.1 Теория

Иногда при взятии производных по вектору или от вектор-функций удобно оперировать матричными операциями. Это сокращает запись и упрощает вывод формул. Введём следующие определения:

- При отображении вектора в число $f(x) : \mathbb{R}^n \to \mathbb{R}$
  $\nabla_x f(x) = [\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}]^T$.
- При отображении матрицы в число $f(A) : \mathbb{R}^{n\times m} \to \mathbb{R}$
  $\nabla_A f(A) = (\frac{\partial f}{\partial A_{ij}})_{i,j=1}^{n,m}$.

Мы хотим оценить, как функция изменяется по каждому из аргументов по отдельности. Поэтому производной функции по вектору будет вектор, по матрице — матрица.

Полезные свойства:

1) $d(XY) = dX \cdot Y + X \cdot dY$
2) Если A - матрица константа, то $dA = 0$
3) $d(X') = dX'$
4) $d \det X = \det X tr(X^{-1}dX)'$

1.2 Семинар

Задача 1 Пусть $a \in \mathbb{R}^n$ – вектор параметров, а $x \in \mathbb{R}^n$ – вектор переменных. Необходимо найти производную их скалярного произведения по вектору переменных $\nabla_x a^T x$.

Задача 2 Пусть $A \in \mathbb{R}^{n\times n}$. Необходимо найти $\nabla_A detA$.

Задача 3 Пусть $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n}$. Необходимо найти $\nabla_A tr(AB)$.

Задача 4 Пусть $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times m}$, $y \in \mathbb{R}^m$. Необходимо найти $\nabla_A x^T Ay$.

1.3 Домашнее задание

Задача 5 Пусть t – скалярная переменная, r, s – векторные переменные, R, S – матричные переменные. Кроме того, a, b – векторы констант, A, B – матрицы констант. Применив базовые правила дифференцирования найдите:

1. $d(ARB)$
2. $d(r'r)$

3. $d(r'Ar)$
4. $d(R^{-1})$, воспользовавшись тем, что $R^{-1}R = I$
5. $d(cos(r'r))$
6. $d(r'Ar/r'r)$

Задача 6 Пусть $A \in \mathbb{R}^{n \times n}$. Необходимо найти $\nabla_x x^T Ax$

2 Линейные классификаторы часть 1

2.1 Семинар

Задача 7 Линейный классификатор выдал следующие значения для объектов из набора данных: $b_i = \mathbb{P}(y_i = 1|x_i)$.

y_i	b_i
1	0.7
-1	0.2
-1	0.3
-1	0.8
1	0.6

Необходимо:
1. Построить ROC-кривую
2. Найти площадь под ROC-кривой и индекс Джини
3. Построить PR-кривую (кривая точность-полнота)
4. Найти площадь под PR-кривой

Задача 8 Дан алгоритм классификации, который выдаёт вероятность принадлежности объекта к положительному классу $b_i = \mathbb{P}(y_i = 1|x_i)$. Также есть набор данных с объектами двух типов: 100 китов и 900 муравьев. В качестве признака алгоритм использует количество глаз у объекта (у китов и муравьёв 2 глаза). После применения алгоритма к набору данных для каждого объекта было получено число $b_i = f(x_i) \in [0; 1]$, оценка вероятности того, что наблюдение является китом.

Решите задачу минимизации эмпирической функции риска и найдите все b_i для функций потерь:

1. $L(y_i, b_i) = (y_i - b_i)^2$, если для муравьёв $y_i = 0$;

2. $L(y_i, b_i) = |y_i - b_i|$, если для муравьёв $y_i = 0$;

2.2 Домашнее задание

Задача 9 Рассмотрим плоскость в $\mathbb{R}^3$, задаваемую уравнением $5x_1+6x_2-7x_3+10 = 0$ и две точки, $A = (2, 1, 4)$ и $B = (4, 0, 4)$.
1. Найдите любой вектор, перпендикулярный плоскости;
2. Правда ли, что отрезок AB пересекает плоскость?
3. Найдите длину отрезка AB;
4. Не находя расстояние от точек до плоскости, определите, во сколько раз точка A дальше от плоскости, чем точка B;

5. Найдите расстояние от точки A до плоскости.

Задача 10 Закончите предложения:
1. accuracy — это доля правильных ответов...
2. точность (precision) — это доля правильных ответов...
3. полнота (recall) — это доля правильных ответов...
4. TPR — это доля правильных ответов...

Задача 11 Алгоритм бинарной классификации выдаёт оценки вероятности $b_i = \mathbb{P}(y_i = 1|x_i)$. Всего проведено 10000 наблюдений. Если ранжировать их по возрастанию b_i , то окажется что наблюдения с $y_i = 1$ занимают ровно места с 5501 по 5600. Найдите площадь по ROC-кривой и площадь под PR-кривой.

Задача 12 Для условия задачи 2.2. минимизируйте эмпирическую функцию риска и найдите все b_i для функций потерь:

1. $L(y_i, b_i) = \begin{cases} \text{-log b}_i\text{, если } y_i = 1 \\ \text{-log (1 - b}_i\text{), иначе} \end{cases}$;

2. $L(y_i, b_i) = \begin{cases} 1/\text{b}_i\text{, если } y_i = 1 \\ 1/(1\text{ - b}_i\text{), иначе} \end{cases}$;

3 Линейные классификаторы часть 2

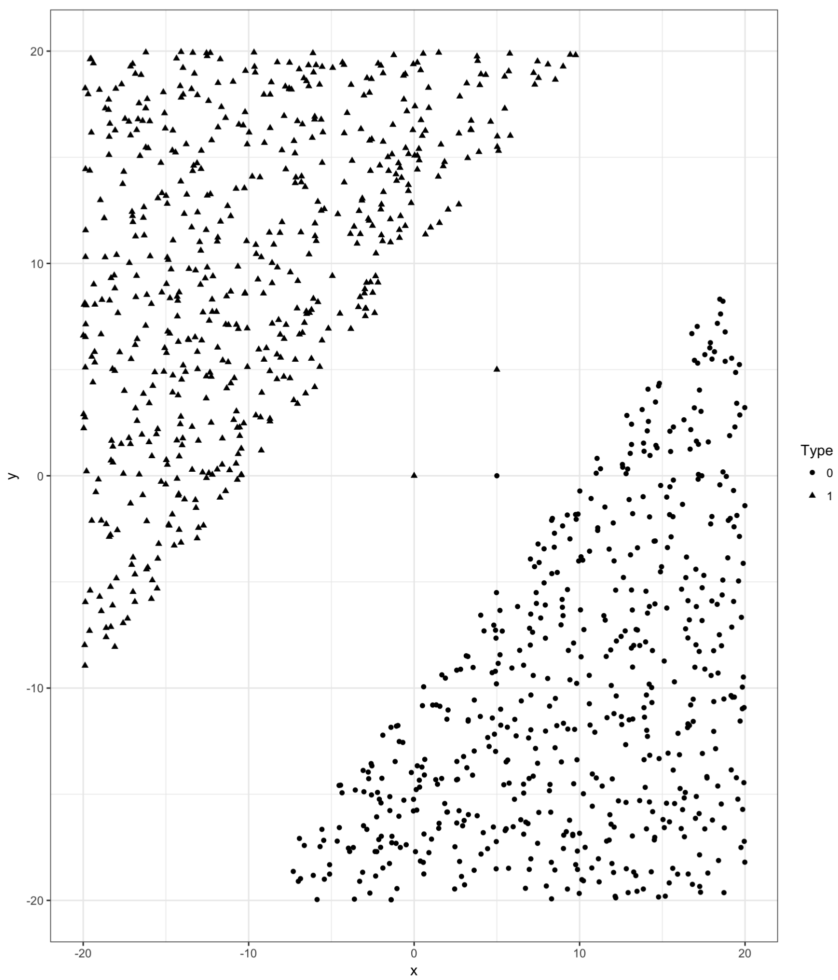
3.1 Семинар

Задача 13 Построить персептон, реализующий логическое ИЛИ.

Задача 14 На плоскости имеются точки двух цветов. Красные: $(1, 1), (1, -1)$ и синие: $(-1, 1), (-1, -1)$.

1. Найдите разделяющую полосу методом опорных векторов при разных C;
2. Укажите опорные вектора.

Задача 15 По картинке качественно решите задачу разделения точек:



Целевая функция имеет вид:

$$\min_{w,w_0} \frac{1}{2}w'w + C\sum_{i=1}^{n}\epsilon_i$$

Уравнение разделяющей поверхности: $w'x = w_0$, уравнения краёв полосы: $w'x = w_0 + 1$ и $w'x = w_0 - 1$. Нарушителями считаются наблюдения, которые попали на нейтральную полосу или на чужую территорию. Здесь $\epsilon_i = |w| \cdot d_i$, где d_i - заступ наблюдения за черту.

1. Как пройдёт разделяющая полоса при $C = 1$? Найдите w, w_0, величины штрафов ϵ_i.
2. Как пройдёт разделяющая полоса при $C = +\infty$? Найдите w, w_0, величины штрафов ϵ_i.

Задача 16 Дана таблица сопряженности x и y:

	$y_i = 1$	$y_i = 0$
$x_i = 1$	12	36
$x_i = 0$	32	20

Будем использовать логистическую регрессию с константой для прогнозирования y с помощью x.

1. Какие оценки коэффициентов мы получим?
2. Какой прогноз вероятности $y = 1$ при значении признака $x = 0$ даёт логистическая модель? Как это число можно посчитать без рассчитывания коэффициентов?

Задача 17 Показать, что из формулы логистической регрессии

$$p(y = +1|x) = \frac{1}{1 + e^{-w^T x}}$$

следует, что (w, x) - это логарифм отношения шансов (log-odds).

Задача 18 Показать, что квадратичная функция потерь

$$L(y, z) = ([y = +1] - z)^2$$

позволяет предсказывать корректные вероятности.