Сборник задач по курсу "Машинное обучение"

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

12 ноября 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 1.1. 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 1.2. Пусть $A \in \mathbb{R}^{n\times n}$. Необходимо найти $\nabla_A detA$.

Задача 1.3. Пусть $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n}$. Необходимо найти $\nabla_A tr(AB)$.

Задача 1.4. 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 1.5. 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 1.6. Пусть $A \in \mathbb{R}^{n \times n}$. Необходимо найти $\nabla_x x^T Ax$

2 Линейные классификаторы часть 1

2.1 Семинар

Задача 2.1. 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 2.2. 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 2.4. 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 2.5. 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 2.6. 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 2.7. 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

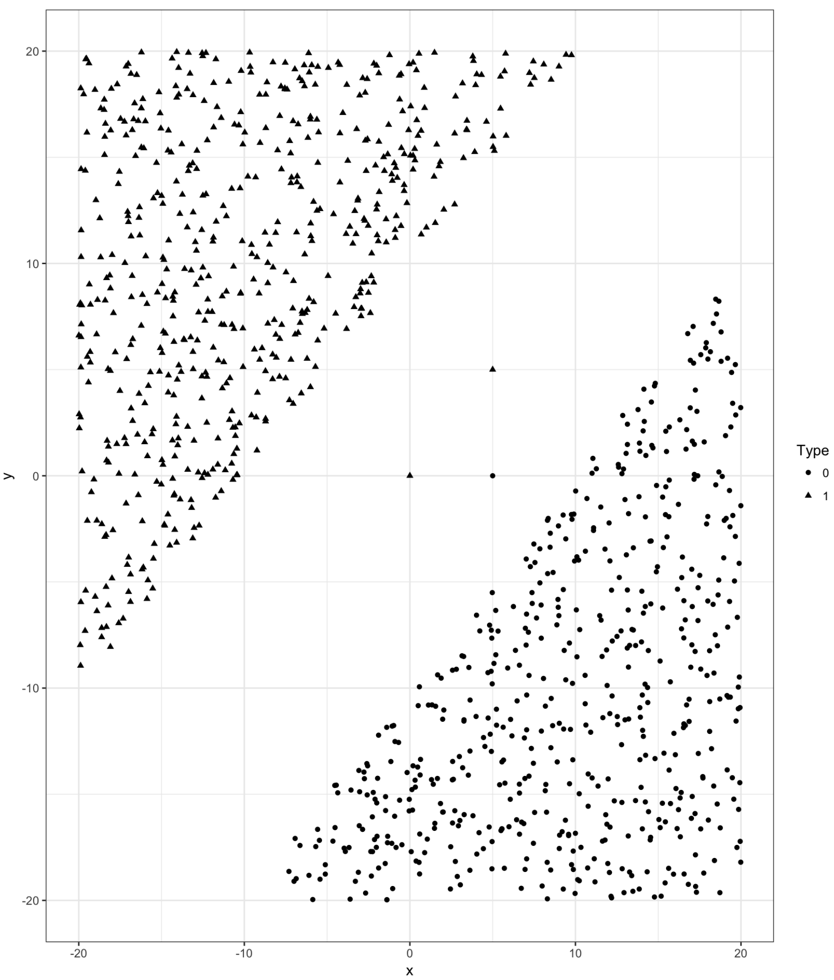
3.1 Семинар

Задача 3.1. Построить персептон, реализующий логическое ИЛИ.

Задача 3.2. 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 3.3. 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 3.4. 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 3.5. 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 3.6. Показать, что квадратичная функция потерь

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

позволяет предсказывать корректные вероятности.

3.2 Домашнее задание

Задача 3.7. Построить персептон, реализующий логическое НЕ.

Задача 3.8.* Построить двухслойный персептон, реализующий XOR.

Задача 3.9. На плоскости имеются точки двух цветов. Красные: $(1, 1)$, $(1, -1)$ и синие: $(-1, 1)$, $(-1, -1)$ и $(2, 0)$.

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

Задача 3.10. Покажите, что абсолютная функция потерь

$$L(y, z) = |[y = +1] - z|, \ z \in [0; 1]$$

не позволяет предсказывать корректные вероятности.

Задача 3.11. Рассмотрим целевую функцию логистической регрессии с константой

$$Q(w) = \frac{1}{l} \sum_{i=1}^{n} L(y_i, b_i),$$

где $b_i = 1 \ / \ (1 + exp(-\langle w, x_i \rangle))$ и $L(y_i, b_i) = \begin{cases} -log \ b_i, \text{ если } y_i = 1 \\ -log(1 - b_i), \text{ иначе} \end{cases}$

Найдите градиент *Q(w)*.

Задача 3.12. Оценка логистической модели для прогнозирования y_i, в зависимости от x_i и z_i: $ln \ odds_i = 2 + 0.3x_i - 0.5z_i$. Оцените вероятность того, что $y_i = 1$ для $x = 15, z = 3.5$.

4 Ядра

4.1 Семинар

Задание 4.1. Переход из двумерного пространства в расширяющее задан функцией

$$f : (x_1, x_2) \longrightarrow (1, x_1, x_2, 3x_1x_2, 2x_1^2, 4x_2^2).$$

Найдите соответствующую ядерную функцию.

Задание 4.2. Ядерная функция имеет вид

$$K(x, y) = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2.$$

Как может выглядеть функция $f : R^2 \longrightarrow R^3$ переводящие исходные векторы в расширенное пространство?

Задание 4.3. Является ли функция $K(x, z) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{otherwise} \end{cases};$ ядром?

Задание 4.4. Докажите, что произведение ядер является ядром.

Задание 4.5. Докажите, что RBF-ядро это ядро.

4.2 Домашнее задание

Задание 4.6. Ядерная функция, скалярное произведение в расширяющем пространстве, имеет вид $K(a, b) = exp(-|a - b|^2)$. Имеются вектора $a = (1, 1, 1)$ и $b = (1, 2, 0)$.

Найдите длину векторов и косинус угла между ними в исходном и расширяющем (спрямляющем) пространстве.

Задание 4.7. Рассмотрим два вектора, $v_1 = (1, 1, 2)$ и $v_2 = (1, 1, 1)$. Переход в спрямляющее пространство осуществляется с помощью гауссовской ядерной функции с параметром γ, $k(v, v') = exp(-\gamma|v - v'|^2)$.

1. Как от γ зависят длины векторов в спрямляющем пространстве?
2. Как от γ зависит угол между векторами в спрямляющем пространстве?

Задание 4.8. Имеются три наблюдения A, B и C:

	x	y
A	1	-2
B	2	1
C	3	0

1. Найдите расстояние AB и косинус угла ABC.
2. Найдите расстояние AB и косинус угла ABC в расширенном пространстве с помощью гауссовского ядра с $K(x, x') = exp(-|x - x'|^2)$.
3. Найдите расстояние AB и косинус угла ABC в расширенном пространстве с помощью полиномиального ядра второй степени.

Задание 4.9. Является ли функция $K(x, z)$ ядром?

1. $K(x, z) = \begin{cases} 0 & \text{if } x = z \\ 1 & \text{otherwise} \end{cases};$

2. $K(x, z) = cos(x^T x) sin(z^T z);$

3. $K(x, z) = sin(x^T z);$

5 Снижение размерности

5.1 Семинар

Задание 5.1. Найдите SVD-разложение матриц.

1. X = $\begin{pmatrix} 0 & -7 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$

2. X = $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

Задание 5.2. Дано сингулярное разложение матрицы X

$$X = U \cdot \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \cdot V'$$

Если возможно найдите сингулярное разложение для
1. $X', 10X$;
2. $X'X, XX'$;

Задание 5.3. Найдите прямую, у которой сумма квадратов расстояний до точек $(0,0), (1,1), (2,1)$ будет минимальной. Чему равна при этом доля объяснённого разброса точек?

Задание 5.4. Есть две переменных, $x = (1,0,0,3)', z = (3,2,0,3)'$. Найдите первую и вторую главные компоненты (то есть, собственные векторы матрицы $X^T X$).

5.2 Домашнее задание

Задание 5.5. Найдите SVD-разложение матриц.

1. X = $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$

2. X = $\begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 2 & 2 \end{pmatrix}$

3. X = $\begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$

Задание 5.6. Дано сингулярное разложение матрицы X

$$X = U \cdot \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \cdot V'$$

Если возможно, найдите сингулярное разложение для
1. $(X'X)^4, (XX')^{-1}, 5I + X'X$;
2. $X'(XX')^{-1}X$;
3. $X(X'X)^{-1}X'$;

Задание 5.7. Известна матрица выборочных ковариаций трёх переменных. Для удобства будем считать, что переменные уже центрированы.

$$\begin{pmatrix} 11 & 2 & -8 \\ 2 & 2 & 10 \\ -8 & 10 & 5 \end{pmatrix}$$

Найдите первую и вторую главные компоненты. Какую долю дисперсии они объясняют?