1 Дифференциал

Минитеория:

1. $d(XY) = dX \cdot Y + X \cdot dY$
2. $dA = 0$
3. $d(X') = dX'$
4. $d\det X = \operatorname{tr}(...)$

1.1 Вспомним дифференциал :)

1. Известно, что $f(x) = x^2 + 3x$. Найдите $f'(x)$ и df. Чему равен dx в точке $x = 5$ при $dx = 0.1$?
2. Известно, что $f(x_1, x_2) = x_1^2 + 3x_1x_2^3$. Найдите df. Чему равен df в точке $x_1 = -2$, $x_2 = 1$ при $dx_1 = 0.1$ и $dx_2 = -0.1$?
3. Известно, что $F = \begin{pmatrix} 5 & 6x_1 \\ x_1x_2 & x_1^2x_2 \end{pmatrix}$. Найдите dF.
4. Известно, что $F = \begin{pmatrix} 7 & 8 & 9 \\ 2 & -1 & -2 \end{pmatrix}$. Найдите dF.
5. Матрица F имеет размер 2×2, в строке i столбце j у неё находится элемент f_{ij}. Выпишите выражение $\operatorname{tr}(F'dF)$ в явном виде без матриц.

1.2 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 $R^{-1} \cdot R = I$;
5. $d\cos(r'r)$;
6. $d(r'Ar/r'r)$.

1.3 В методе наименьших квадратов минимизируется функция

$$Q(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

1. Найдите $dQ(\hat{\beta})$ и $d^2Q(\hat{\beta})$;
2. Выпишите условия первого порядка для задачи МНК;
3. Выразите $\hat{\beta}$ предполагая, что $X'X$ обратима.

1.4 В методе LASSO минимизируется функция

$$Q(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) + \lambda\hat{\beta}'\hat{\beta},$$

где λ — положительный параметр, штрафующий функцию за слишком большие значения $\hat{\beta}$.

1. Найдите $dQ(\hat{\beta})$ и $d^2Q(\hat{\beta})$;
2. Выпишите условия первого порядка для задачи LASSO;
3. Выразите $\hat{\beta}$.

1.5 Пусть A и B — матрицы одного размера.

1. Докажите, что сумму $\sum_{ij} A_{ij}B_{ij}$ можно представить в виде $\operatorname{tr}(A'B)$.
2. Докажите, что $\operatorname{tr}(A'B) = \operatorname{tr}(AB') = \operatorname{tr}(B'A) = \operatorname{tr}(BA')$.

1.6 Выведите формулу для $d \det X$.

1.7 Пусть x_i — вектор-столбец $k \times 1$, y_i — скаляр, равный $+1$ или -1, $\hat{\beta}$ — вектор-столбец размера $k \times 1$. Рассмотрим функцию

$$Q(\hat{\beta}) = \sum_{i=1}^{n} \ln(1 + \exp(-y_i x_i' \hat{\beta})) + \lambda \hat{\beta}' \hat{\beta}$$

1. Найдите dQ;
2. Найдите вектор-столбец $\operatorname{grad} Q$.

2 Линейная регрессия

2.1 Рассмотрим задачу линейной регресии

$$Q(w) = (y - Xw)^T (y - Xw) \to \min_w .$$

1. Найдите $dQ(w)$ и $d^2Q(w)$.
2. Выведите формулу для оптимального w.
3. Выведите формулу для матрицы-шляпницы (hat-matrix), связывающей вектор фактических y и вектор прогнозов $\hat{y} = H \cdot y$.

2.2 Рассмотрим задачу регрессии с одним признаком и без константы, $\hat{y}_i = w \cdot x_i$. Решите в явном виде задачи МНК со штрафом:

1. $Q(w) = (y - \hat{y})^T (y - \hat{y}) + \lambda w^2$;
2. $Q(w) = (y - \hat{y})^T (y - \hat{y}) + \lambda |w|$;

2.3 Храбрая и торопливая исследовательница Мишель хочет решить задачу линейной регрессии по n наблюдениям с вектором y и матрицей признаков X. Сначала исследовательница Мишель так торопилась, что совсем забыла последнее наблюдение и оценила задачу с более коротким вектором y^- и матрицей X^-, где не хватает последней строки. Затем Мишель взяла правильную матрицу X, но неправильный вектор y^*, в котором она вместо фактического последнего наблюдения вектора y вписала его прогноз, полученный с помощью регрессии с y^{-1} и X^-.

1. Как связаны $\hat{y}_n^-$ и $\hat{y}_n^*$ (прогнозы для последнего наблюдения полученные по модели без последнего наблюдения и модели с неверным последним наблюдением)?
2. Как выглядит вектор, равный разнице $y - y^*$?

3. Какие величины находятся в векторе $H \cdot (y - y^*)$? Чему равна последняя, n-ая, компонента этого вектора? Выразите её через H_{nn} и ошибку прогноза последнего наблюдения по модели без последнего наблюдения, $y_n - \hat{y}_n^-$.
4. Как связаны между собой ошибка прогноза n-го наблюдения по полной модели, ошибка прогноза n-го наблюдения по модели без последнего наблюдения и H_{nn}?
5. Как быстро провести кросс-валидацию с выкидыванием одного наблюдения для задачи линейной регрессии?

3 Линейные классификаторы

3.1 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.

3.2 Рассмотрим простейший персептрон с константой, единственным входом x_1 и пороговой функцией активации. Подберите веса так, чтобы персептрон реализовывал логическое отрицание (в ответ на 0 выдавал 1, и наоборот).

3.3 Рассмотрим простейший персептрон с константой, двумя входами x_1, x_2 и пороговой функцией активации.

Здесь ассистенты нарисуют в tikz картинку, достойную стоять вместо Джоконды в Лувре

1. Подберите веса так, чтобы персептрон реализовывал логическое ИЛИ (OR).
2. Подберите веса так, чтобы персептрон реализовывал логическое И (AND).
3. Докажите, что веса невозможно подобрать так, чтобы персептрон реализовывал исключающее логическое ИЛИ (XOR).
4. Добавьте персептрону вход $x_3 = x_1 \cdot x_2$ и подберите веса так, чтобы персептрон реализовывал XOR.
5. Реализуйте XOR с помощью трёх персептронов с двумя входами и константой. Укажите веса и схему их взаимосвязей.

3.4 В коробке завалялось три персептрона, у каждого два входа с константой и пороговая функция активации. Реализуйте с их помощью функцию

$$y = \begin{cases} 1, & \text{если } x_2 \geqslant |x_1 - 3| + 2; \\ 0, & \text{иначе} \end{cases} .$$

3.5 Рассмотрим следующий набор данных:

x_i	z_i	y_i
-1	-1	0
1	-1	0
-1	1	0
1	1	0
0	2	1
2	0	1
0	-2	1
-2	0	1

1. Существует ли перспетрон с константой, двумя входами и пороговой функцией активации, способный идеально классифицировать y_i на данной выборке? А хватит ли двух таких персептронов? А может хватит трёх?
2. Введите такое преобразование исходных признаков $h_i = h(x_i, z_i)$, при котором с идеальной классификацией y_i справился бы даже персептрон с одним входом, константой и пороговой функцией активации.

3.6 Бандерлог из Лога[1] ведёт блог, любит считать логарифмы и оценивать логистические регрессии. С помощью нового алгоритма Бандерлог решил задачу классификации по трём наблюдениям и получил $b_i = \hat{\mathbb{P}}(y_i = 1|x_i)$.

y_i	b_i
1	0.7
-1	0.2
-1	0.3

1. Постройте ROC-кривую.
2. Найдите площадь под ROC-кривой и индекс Джини.
3. Постройте PR-кривую (кривая точность-полнота).
4. Найдите площадь под PR-кривой.
5. Как по-английски будет «бревно»?

3.7 Классификатор Бандерлога имеет вид

$$a_i = \begin{cases} 1, \text{ если } b_i > t; \\ -1, \text{ иначе.} \end{cases}$$

Докажите, что площадь под ROC-кривой равна вероятности того, случайно выбранный положительный объект окажется позже случайно выбранного отрицательного объекта, если объекты ранжированы по возрастанию величины b_i.

3.8 Все средние издалека выглядят одинаково, среднее $= f^{-1}(0.5f(x_1) + 0.5f(x_2))$. Например, у среднего арифметического $f(t) = t$, у среднего гармонического $f(t) = 1/t$.

1. Какая f используется для среднего геометрического?

Для измерения качества бинарной классификации Ара использует среднее арифметическое точности и полноты, Гена — среднее геометрическое, а Гарик — среднее гармоническое.

[1] деревня в Кадуйском районе Вологодской области

2. У кого будут выходить самые «качественные» и самые «некачественные» прогнозы?

3.9 Бандерлог начинает все определения со слов «это доля правильных ответов»:

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…

Закончите определения Бандерлога так, чтобы они были, хм, правильными.

3.10 Алгоритм бинарной классификации, придуманный Бандерлогом, выдаёт оценки вероятности $b_i = \hat{\mathbb{P}}(y_i = 1|x_i)$. Всего у Бандерлога 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.

3.11 Бандерлог собрал выборку из 900 муравьёв и 100 китов. Переменная y_i равна 1 для китов. Бандерлог хочет, чтобы его алгоритм классификации выдавал для каждого наблюдения число $b_i = f(x_i) \in [0; 1]$, оценку вероятности того, что наблюдение является китом. В качестве признака Бандерлог использует количество глаз, не задумавшись о том, что оно равно двум и для муравьёв, и для китов.

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 $y_i = 0$;
3. $L(y_i, b_i) = \begin{cases} -\log b_i, & \text{если } y_i = 1 \\ -\log(1 - b_i), & \text{иначе.} \end{cases}$;
4. $L(y_i, b_i) = \begin{cases} 1/b_i, & \text{если } y_i = 1 \\ 1/(1 - b_i), & \text{иначе.} \end{cases}$;

3.12 Бандерлог утверждает, что открыл новую верхнюю границу для пороговой функции потерь, $\tilde{L}(M_i) = 1 + \frac{1}{\pi} \cdot \arctan(-x_i)$, где $M_i = y_i \cdot \langle w, x_i \rangle$. Прав ли бандерлог?

3.13 Бандерлог из Лога оценил логистическую регрессию по четырём наблюдениям и одному признаку с константой, получил $b_i = \hat{\mathbb{P}}(y_i = 1|x_i)$, но потерял последнее наблюдение:

y_i	b_i
1	0.7
-1	0.2
-1	0.3
?	?

1. Выпишите функцию потерь для задачи логистической регрессии.
2. Выпишите условие первого порядка по коэффициенту перед константой.
3. Помогите бандерлогу восстановить пропущенные значения!

3.14 У Бандерлога три наблюдения, первое наблюдение — кит, остальные — муравьи. Киты кодируются $y_i = 1$, муравьи — $y_i = -1$. На этот раз Бандерлог, чтобы быть уверенным, что x_i различаются, сам лично определил $x_i = i$. После этого Бандерлог оценивает логистическую регрессию с константой.

1. Выпишите эмпирическую функцию риска, которую минимизирует Бандерлог;
2. При каких оценках коэффициентов логистической регрессии эта функция достигает своего минимума?

3.15 Рассмотрим целевую функцию логистической регрессии с константой

$$Q(w) = \frac{1}{\ell}\sum L(y_i, b_i),$$

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

1. Найдите $dQ(w)$ и $d^2Q(w)$;
2. Найдите $dQ(0)$ и $d^2Q(0)$;
3. Выпишите квадратичную аппроксимацию для $Q(w)$ в окрестности $w = 0$;
4. С какой задачей совпадает задача минимизации квадратичной аппроксимации?

3.16 Винни-Пух знает, что мёд бывает правильный, $honey_i = 1$, и неправильный, $honey_i = 0$. Пчёлы также бывают правильные, $bee_i = 1$, и неправильные, $bee_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$honey_i = 1$	$honey_i = 0$
$bee_i = 1$	12	36
$bee_i = 0$	32	20

Винни-Пух использует логистическую регрессию с константой для прогнозирования правильности мёда с помощью правильности пчёл.

1. Какие оценки коэффициентов получит Винни-Пух?
2. Какой прогноз вероятности правильности мёда при встрече с неправильными пчёлами даёт логистическая модель? Как это число можно посчитать без рассчитывания коэффициентов?

3.17 Винни-Пух оценил логистическую регрессию для прогнозирования правильности мёда от высоты дерева (м) x_i и удалённости от дома (км) z_i: $\ln odds_i = 2 + 0.3x_i - 0.5z_i$.

1. Оцените вероятность того, что $y_i = 1$ для $x = 15$, $z = 3.5$.
2. Оцените предельный эффект увеличения x на единицу на вероятность того, что $y_i = 1$ для $x = 15$, $z = 3.5$.
3. При каком значении x предельный эффект увеличения x на единицу в точке $z = 3.5$ будет максимальным?

4 Матрицы

4.1 Известна матрица X,

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

1. Найдите QR-разложение матрицы $X'X$;
2. Найдите QR-разложение матрицы XX';
3. Найдите спектральное разложение матрицы $X'X$;
4. Найдите спектральное разложение матрицы XX';
5. Найдите сингулярное разложение (SVD) матрицы X;

4.2 Объясните геометрический смысл QR, SVD и спектрального разложений.

4.3 Бандрелог выполнил SVD-разложение матрицы регрессоров X. Помогите Бандерлогу поскорее найти формулу для матрицы-шляпницы H, которая проецирует y на пространство столбцов матрицы X, $\hat{y} = Hy$.

4.4 Бандрелог выполнил QR-разложение матрицы регрессоров X. Помогите Бандерлогу поскорее найти формулу для матрицы-шляпницы H, которая проецирует y на пространство столбцов матрицы X, $\hat{y} = Hy$.

5 Метод опорных векторов

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

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

5.2 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.

5.3 Эконометресса Авдотья решила использовать метод опорных векторов с гауссовским ядром с параметром $\sigma = 1$ и штрафным коэффициентом $C = 1$. Соответственно, она минимизировала целевую функцию

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

где разделяющая плоскость задаётся $w'x - w_0 = 0$, а ξ_i — размеры «заступа» за разделяющую полосу.

Затем Автдотья подумала, что неплохо бы выбрать наилучшие C и σ. Ей лень было использовать кросс-валидацию, поэтому Авдотья минимизировала данную функцию по $C \geqslant 0$ и $\sigma \geqslant 0$. Какие значения она получила?

5.4 Задан вектор $w = (2, 3)$ и число $w_0 = 7$.

1. Нарисуйте прямые $\langle w, x\rangle = w_0$, $\langle w, x\rangle = w_0 + 1$, $\langle w, x\rangle = w_0 - 1$.

2. Найдите ширину полосы между $\langle w, x\rangle = w_0 + 1$ и $\langle w, x\rangle = w_0 - 1$.

3. Найдите расстояние от точки $(5, 6)$ до прямой $\langle w, x\rangle = w_0 - 1$.

5.5 Заданы две прямые, l_0: $x^{(1)} + 3x^{(2)} = 9$ и l_1: $x^{(1)} + 3x^{(2)} = 13$. Найдите подходящий вектор w и число w_0 так, чтобы прямая l_0 записывалась как $\langle w, x\rangle = w_0 - 1$, а прямая l_1 как $\langle w, x\rangle = w_0 + 1$.

5.6 Даны наблюдения

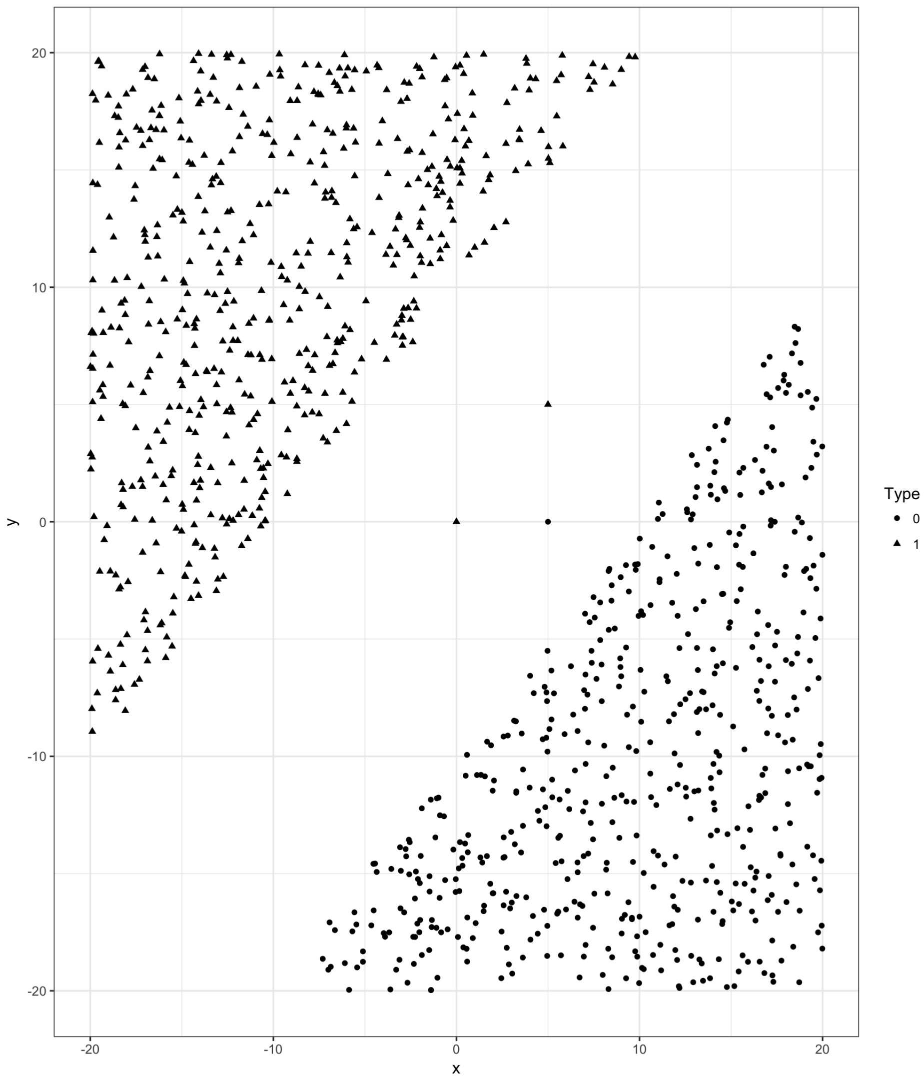
$x^{(1)}$	$x^{(2)}$	y
1	0	0
2	0	0
0	3	1
0	4	1

1. Нарисуйте разделяющую полосу наибольшей ширины.

2. Решите задачу оптимизации
$$\min_{w, w_0} \frac{1}{2}\langle w, w\rangle$$
при ограничении: для $y_i = 1$ выполнено условие $\langle w, x\rangle \geqslant w_0 + 1$, а для $y_i = 0$ выполнено условие $\langle w, x\rangle \leqslant w_0 - 1$.

3. Для точки $x = (x^{(1)}, x^{(2)}) = (1, 1)$ найдите значение $\langle w, x\rangle - w_0$ и постройте прогноз $\hat{y}$.

5.7 По картинке качественно решите задачу разделения точек:



Целевая функция имеет вид:

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

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 $\xi_i = |w| \cdot d_i$, где d_i — длина «заступ» наблюдения за черту «своих».

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

5.8 ююю

6 Ядра к бою!

6.1 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 пространстве.

6.2 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?

6.3 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.

6.4 Переход из двумерного пространства в расширяющее задан функцией

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

Найдите соответствующую ядерную функцию.

6.5 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 $f : \mathbb{R}^2 \to \mathbb{R}^3$ переводящие исходные векторы в расширенное пространство?

6.6 Является ли функция $K(x,z)$ ядром?

1. $K(x,z) = \begin{cases} 1, & \text{if } x = z; \\ 0, & \text{otherwise} \end{cases}$;
2. $K(x,z) = \begin{cases} 0, & \text{if } x = z; \\ 1, & \text{otherwise} \end{cases}$;
3. $K(x,z) = \sin(x^T z)$;
4. $K(x,z) = \cos(x^T x)\sin(z^T z)$;

6.7 Пусть x и z — строки символов, возможно разной длины. Рассмотрим две функции. Функция $K_1(x,z)$ равна единице, если строки x и z совпадают. Функция $K_2(x,z)$ — число совпадающих подстрок. Функция K_3 — произведение количеств букв «а» в обеих словах.

1. Найдите K_1(«мама», «ам») и K_2(«мама», «ам»), K_3(«мама», «ам»)
2. Является ли функция K_1 ядром?
3. Является ли функция K_2 ядром?
4. Является ли функция K_3 ядром?

6.8 На прямой аллее растёт три дуба. Находятся в точках с координатами $x_1 = 1$, $x_2 = 2$ и $x_3 = -3$. Исследователь Винни-Пух проверил и выяснил, что на втором Дубе водятся правильные пчёлы, а на остальных — неправильные.

1. Являются ли пчёлы линейно разделимыми в пространстве исходной аллеи?
2. Помогите Винни-Пуху выписать прямую задачу метода опорных векторов в пространстве исходной аллеи;
3. Помогите Винни-Пуху выписать двойственную задачу метода опорных векторов в пространстве исходной аллеи;
4. Помогите Винни-Пуху выписать двойственную задачу метода опорных векторов в бесконечномерном пространстве с ядерной функцией $K(x,z) = \exp(-(x-z)^2)$; Являются ли точки в нём линейно разделимыми?
5. Помогите Винни-Пуху выписать прямую и двойственную задачу метода опорных векторов в спрямляющем пространстве с ядерной функцией $K(x,z) = (xz+1)^2$; Являются ли точки в нём линейно разделимыми?

7 Двойственные задачи

7.1 Выпишите двойственную задачу для минимизации $x_1^2 + x_2^2 + x_3^2$ при ограничении $2x_1 + 3x_2 + 5x_3 = 10$.

7.2 Выпишите двойственную задачу для $x_1 + 2x_2 + 3x_3 \to \max$ при ограничениях $x_1 + x_2 + x_3 \leqslant 10$, $2x_1 + x_2 + x_3 \leqslant 10$, все $x_i \geqslant 0$.

7.3 Выпишите двойственную задачу для максимизации $1/x_1 + 2/x_2$ при ограничении $2x_1 + 3x_2 = 10$ и $x_1 \in [1;10]$, $x_2 \in [2;6]$.

7.4 Выпишите двойственную задачу для минимизации $f(x) = \frac{1}{2}x'Hx + g'x$ при ограничении $A'x = b$.

7.5 Выпишите двойственную задачу для минимизации $f(x) = \frac{1}{2}x'Hx + g'x$ при ограничении $A'x \leqslant b$.

7.6 Выпишите прямую и двойственную задачу для метода опорных векторов в исходном пространстве.

7.7 Выпишите прямую и двойственную задачу для метода опорных векторов в спрямляющем пространстве с использованием ядра $K(.,.)$.

8 Метод главных компонент

8.1 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?

8.2 Есть две переменных, $x = (1,0,0,3)'$, $z = (3,2,0,3)'$. Найдите первую и вторую главные компоненты.

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

$$\begin{pmatrix} 4 & 1 & -1 \\ 1 & 5 & 0 \\ -1 & 0 & 9 \end{pmatrix}$$

1. Выразите первую и вторую главные компоненты через три исходных переменных.
2. Выразите первую и вторую главные компоненты, через три исходных переменных, если перед методом главных компонент переменные необходимо стандартизировать.

8.4 Пионеры, Крокодил Гена и Чебурашка собирали металлолом несколько дней подряд. В распоряжение иностранной шпионки, гражданки Шапокляк, попали ежедневные данные по количеству собранного металлолома: вектор g — для Крокодила Гены, вектор h — для Чебурашки и вектор x — для Пионеров. Гена и Чебурашка собирали вместе, поэтому выборочная корреляция $\mathrm{sCorr}(g,h) = -0.9$. Гена и Чебурашка собирали независимо от Пионеров, поэтому выборочные корреляции $\mathrm{sCorr}(g,x) = 0$, $\mathrm{sCorr}(h,x) = 0$. Если регрессоры g, h и x центрировать и нормировать, то получится матрица $\tilde{X}$.

1. Найдите параметр обусловленности матрицы $(\tilde{X}'\tilde{X})$.
2. Вычислите одну или две главные компоненты (выразите их через вектор-столбцы матрицы. $\tilde{X}$), объясняющие не менее 70% общей выборочной дисперсии регрессоров.
3. Шпионка Шапокляк пытается смоделировать ежедневный выпуск танков, y. Выразите оценки коэффициентов регрессии $y = \beta_1 + \beta_2 g + \beta_3 h + \beta_4 x + \varepsilon$ через оценки коэффициентов регрессии на главные компоненты, объясняющие не менее 70% общей выборочной дисперсии.

9 На природу! В лес! К деревьям!

9.1 Для случайных величин X и Y найдите индекс Джини и энтропию

x	0	1
$\mathbb{P}(X = x)$	0.2	0.8

,

y	0	1	5
$\mathbb{P}(Y = y)$	0.2	0.3	0.5

9.2 Найдите энтропию X, спутанность (perplexity) X, индекс Джини X, если

1. величина X равновероятно принимает значения 1, 7 и 9;
2. величина X равновероятно принимает $k \geqslant 2$ значений;
3. величина X равномерно распределена на отрезке $[0; a]$;
4. величина X нормальна $\mathcal{N}(\mu; \sigma^2)$;

9.3 У Васи была дискретная случайная величина X, принимавшая натуральные значения. Вася решил изменить закон распределения величины X. Он увеличил количество возможных значений величины X в два раза, разделив каждое событие $X = k$ на два равновероятных подсобытия: $X = k - 0.1$ и $X = k + 0.1$. Как при этом изменились энтропия, спутанность (perplexity) и индекс Джини?

9.4 Случайная величина X принимает значение 1 с вероятностью p и значение 0 с вероятностью $1 - p$.

1. Постройте график зависимости индекса Джини и энтропии от p.
2. Являются ли функции монотонными? выпуклыми?
3. При каком p энтропия и индекс Джини будут максимальны?

9.5 Шаман Ыуыуыуыыыыы по прошлым наблюдениям знает, большая охота на мамонта оказывается удачной с вероятностью 0.3. Если племя ждёт от Ыуыуыуыыыыы прогноз охоты, то Ыуыуыуыыыыы поплясав вокруг костра (10 минут) и постуча бубном (16 раз) прогнозирует удачную охоту с вероятностью 0.3 и неудачную с вероятностью 0.7.

1. Какова вероятность того, что Ыуыуыуыыыыы ошибётся?
2. Чему равен индекс Джини для случайной величины равной удаче с вероятностью 0.3 и неудаче с вероятностью 0.7?

9.6 Шаман Ыуыуыуыыыыы заметил по прошлым данным, что в дождливые дни большая охота на мамонта удачна с вероятностью 0.7, а в сухие — с вероятностью 0.1. Поэтому в дождливый день Ыуыуыуыыыыы предскажет удачу с вероятностью 0.7, а в сухой — с вероятностью 0.1. Дождливых дней — 20%.

1. Какова вероятность того, что Ыуыуыуыыыыы ошибётся?
2. Чему равен индекс Джини выборки разделённой на две части: в части А шесть бананов и 14 апельсинов, а в части В — восемь бананов и 72 апельсина?

9.7 Постройте регрессионное дерево для прогнозирования y с помощью x на обучающей выборке:

x_i	0	1	2	3
y_i	5	6	4	100

Критерий деления узла на два — минимизация RSS. Дерево строится до трёх терминальных узлов.

9.8 Постройте регрессионное дерево для прогнозирования y с помощью x на обучающей выборке:

y_i	x_i
100	1
102	2
103	3
50	4
55	5
61	6
70	7

Критерий деления узла на два — минимизация RSS. Узлы делятся до тех пор, пока в узле остаётся больше двух наблюдений.

9.9 Дон-Жуан предпочитает брюнеток. Перед Новым Годом он посчитал, что в записной книжке у него 20 блондинок, 40 брюнеток, две рыжих и восемь шатенок. С Нового Года Дон-Жуан решил перенести все сведения в две записные книжки, в одну — брюнеток, во вторую — остальных.

Как изменились индекс Джини и энтропия в результате такого разбиения?

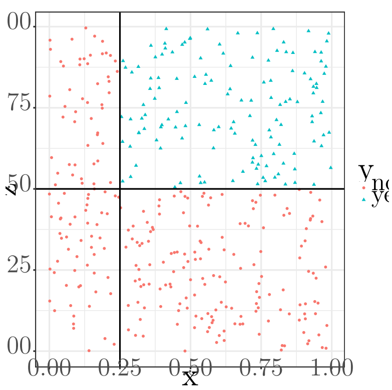
9.10 Машка пять дней подряд гадала на ромашке, а затем выкладывала очередную фотку «Машка с ромашкой» в инстаграмчик. Результат гадания — переменная y_i, количество лайков у фотки — переменная x_i. Постройте классификационное дерево для прогнозирования y_i с помощью x_i на обучающей выборке:

y_i	x_i
плюнет	10
поцелует	11
поцелует	12
к сердцу прижмёт	13
к сердцу прижмёт	14

Дерево строится до идеальной классификации. Критерий деления узла на два — максимальное падение индекса Джини.

9.11 У Винни-Пуха есть 100 песенок (кричалок, вопелок, пыхтелок и сопелок). Каждый день он выбирает и поёт одну из них равновероятно наугад. Одну и ту же песенку он может петь несколько раз. Сколько в среднем песенок оказываются неспетыми за 100 дней?

9.12 По данной диаграмме рассеяния постройте классификационное дерево для зависимой переменной y:



Дерево необходимо построить до идеальной классификации, в качестве критерия деления узла на два используйте минимизацию индекса Джини.

9.13 Рассмотрим обучающую выборку для прогнозирования y с помощью x и z:

y_i	x_i	z_i
y_1	1	2
y_2	1	2
y_3	2	2
y_4	2	1
y_5	2	1
y_6	2	1
y_7	2	1

Будем называть деревья разными, если они выдают разные прогнозы на обучающей выборке. Сколько существует разных классификационных деревьев для данного набора данных?

9.14 Исследовательница Мишель строит классификационное дерево для бинарной переменной y_i. Может ли при разбиении узла на два расти индекс Джини? Энтропия?

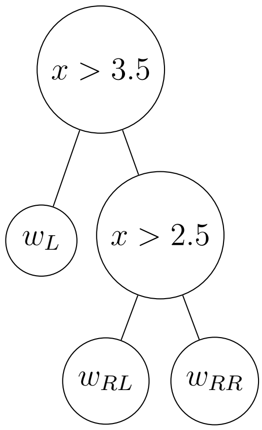
9.15 Приведите примеры наборов данных, для которых индекс Джини равен 0, 0.5 и 0.999.

9.16 Рассмотрим задачу построения классификационного дерева для бинарной переменной y_i. Приведите пример такого набора данных, что никакое разбиения стартового узла на два не снижает индекс Джини, однако двух разбиений достаточно, чтобы снизить индекс Джини до нуля.

9.17 Пятачок собрал данные о визитах Винни-Пуха в гости к Кролику. Здесь x_i — количество съеденого мёда в горшках, а y_i — бинарная переменная, отражающая застревание Винни-Пуха при выходе.

Для построения предиктивной модели Пятачок собирается использовать дерево с заданной структурой:

y_i	x_i
0	1
1	4
1	2
0	3
1	3
0	1



Пятачок использует квадратичную аппроксимацию для логистической функции потерь:

$$Obj(w) = \sum_{i=1}^{n} \left(loss(y_i, 0) + loss'_w(y_i, 0)(w_i - 0) + \frac{1}{2} loss''_{ww}(y_i, 0)(w_i - 0)^2 \right) + \frac{1}{2}\lambda |w|^2.$$

Помогите Очень Маленькому Существу подобрать оптимальные веса (w_i) при $\lambda = 1$.

9.18 Нарисовано дерево: деление 1, справа от первого деления — деление 2. Веса равны w_L, w_{RL}, w_{LL}. Дана выборка.

1. Выпишите в явном виде функцию правдоподобия и логистическую функцию потерь.
2. Оцените w методом максимального правдоподобия.
3. Тут другую функцию потерь написать!
4. Разложите функцию потерь в окрестности $w = (0, 0, 1)$ в ряд Тейлора до второго члена и примерно оцените w.

10 Решения

1.1.

1.2.

1. $A(dR)B$
2. $2r'dr$
3. $r'(A' + A)dr$
4. $R^{-1} \cdot dR \cdot R^{-1}$
5. $-\sin(r'r) \cdot 2r'dr$
6. $\frac{r'(A'+A)dr \cdot r'r - r'Ar2r'dr}{(r'r)^2}$

1.3.

1. $dQ(\hat{\beta}) = 2(y - X\hat{\beta})^T(-X)d\hat{\beta}$, $d^2Q(\hat{\beta}) = 2d\hat{\beta}X^2d\hat{\beta}$
2. $dQ(\hat{\beta}) = 0$
3. $\hat{\beta} = (X^TX)^{-1}X^Ty$

1.4.

1. $dQ(\hat{\beta}) = -2((y - X\hat{\beta})^TX + \lambda\hat{\beta}^T)d\hat{\beta}$, $d^2Q(\hat{\beta}) = 2d\hat{\beta}^T(X^TX - \lambda I)d\hat{\beta}$
2. $dQ(\hat{\beta}) = 0$
3. $\hat{\beta} = (X^TX - \lambda I)^{-1}X^Ty$

1.5.

1.6.

1.7.

2.1.

1. $dQ(w) = 2(Xw - y)^TXdw$, $d^2Q(w) = 2dw^TX^TXdw$
2. $w = (X^TX)^{-1}X^Ty$

3. $\hat{y} = Xw = X(X^TX)^{-1}X^Ty \Rightarrow H = X(X^TX)^{-1}X^T$

2.2.

1. Выпишем $dQ(W)$ и найдём градиент:

$$dQ(w) = 2(Xw - y)^T X dw + 2\lambda w^T dw \Rightarrow \nabla Q(w) = 2X^T(Xw - y) + 2\lambda w$$

   Приравняв градиент к нулю, получим:

$$w = (X^TX + \lambda I)^{-1}X^Ty$$

2.

2.3.

$$y_n - \hat{y}_n = (1 - H_{nn})(y_n - \hat{y}_n^-)$$

3.1.

1. $(5, 6, -7)$
2. Подставим точки A и B в уравнение плоскости:

$$A : 5 \cdot 2 + 6 \cdot 1 - 7 \cdot 4 + 10 = -2$$

$$B : 5 \cdot 4 + 6 \cdot 0 - 7 \cdot 4 + 10 = 2$$

   Точки A и B лежат по разные стороны плоскости, следовательно, отрезок AB пересекает её.

3. $\overrightarrow{AB} = (2, -1, 0)$, $|AB| = \sqrt{4 + 1 + 0} = \sqrt{5}$
4. Расстояние одинаково
5. $\rho = \frac{|5 \cdot 2 + 6 \cdot 1 - 7 \cdot 4 + 10|}{\sqrt{5^2 + 6^2 + 7^2}} = \frac{2}{\sqrt{10}}$

3.2. Например, $w_1 = -2$, $w_0 = 2$, где w_0 — вес при константе.

3.3.

1. Например, $w_1 = 2, w_2 = 2, w_0 = 0$
2. Например, $w_1 = 2, w_2 = 2, w_0 = -2$
3. Можно показать графически: нарисовать на плоскости точки $(0, 0), (1, 1), (1, 0), (0, 1)$, причём для первых двух нейрон должен выдавать ответ 0, а для вторых — 1. Чтобы разделить эти точки, необходимо провести две прямые, в то время как один нейрон проводит только одну.
4. Например, подойдут веса $w_1 = 3, w_2 = 3, w_3 = -5, w_0 = -1$

5. Первый нейрон с весами $w_{11} = 1, w_{12} = 1, w_{10} = 1/2$ и второй нейрон с весами $w_{21} = 1, w_{22} = 1, w_{10} = -1/2$ должны подавать реузльтаты на вход третьему нейрону с весами $w_{31} = 3, w_{32} = -1, w_{30} = -2$

3.4.

3.5.

3.6.

3.7.

3.8.

1. $f(t) = \log(t)$
2. Среднее гармоническое $<$ среднее геометрическое $<$ среднее арифметическое

3.9.

1. accuracy $= \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}$
2. precision $= \frac{\text{TP}}{\text{TP+FP}}$
3. recall $= \frac{\text{TP}}{\text{TP+FN}}$
4. TPR $= \frac{\text{TP}}{\text{TP+FN}}$

3.10.

3.11.

1. Поскольку все признаки одинаковы, то $\forall i \quad b_i = f(x_i) = b$, и функционал ошибки иммет вид:
$$Q(b) = \frac{1}{1000}\sum_{i=1}^{1000} L(y_i, b) = \frac{1}{1000}\left(\sum_{i=1}^{900} b^2 + \sum_{i=1}^{100}(1-b)^2\right) \to \min_b$$
Дифференцируем и находим b:
$$2 \cdot 900 \cdot b - 2 \cdot 100 \cdot (1-b) = 0 \Rightarrow b = 0.1$$

2. Аналогично:
$$Q(b) = \frac{1}{1000}\sum_{i=1}^{1000} |y_i - b| = \frac{1}{1000}(900 \cdot b + 100|1-b|) \to \min_b$$
При $b = 1$, получаем: $Q(1) = 0.9$ При $b < 1$: $Q(b) = \frac{1}{1000}(900b + 100 - 100b) = \frac{1}{1000}(900b + 100) \to \min_b$. Минимум функционала ошибки достигается при $b = 0$ и равен 0.1.

3. Снова выпишем функционал ошибки:

$$Q(b) = \frac{1}{1000}\left(-900\log(1-b) + 100\log b\right) \to \min_b$$

Берём производную и получаем оптимальный b:

$$\frac{1}{1000}\left(\frac{900}{1-b} - \frac{100}{b}\right) = 0 \Rightarrow b = 0.1$$

4. $$Q(b) = \frac{1}{1000}\left(\frac{900}{1-b} - \frac{100}{b}\right) \to \min_b$$

Дифференцируя, получим:

$$\frac{900}{(1-b)^2} = \frac{100}{b^2} \Rightarrow b = 0.25$$

3.12. Нет. Не выполнено $\tilde{L} \geqslant L$ для всех $M \in \mathbb{R}$.

3.13.

3.14.

3.15.

3.16.

3.17. Предельный эффект максимален при максимальной производной $\Lambda'(\hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z)$, то есть при $\hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z = 0$.

4.1.

4.2.

4.3.

4.4.

5.1.

5.2.

5.3. $C = 0$ и $\sigma = +\infty$

5.4.

5.5.

5.6.

5.7.

5.8.

6.1. В исходном пространстве: $|\vec{a}| = \sqrt{3}$, $|\vec{b}| = \sqrt{5}$, $\cos(\vec{a}, \vec{b}) = \sqrt{0.6}$.
В расширяющем пространстве: $|h(\vec{a})| = 1$, $|h(\vec{b})| = 1$, $\cos(h(\vec{a}), h(\vec{b})) = e^{-2}$.

6.2. Длина равна 1 и не зависит от γ. При $\gamma \approx 0$ вектора примерно совпадают, при больших γ вектора примерно ортогональны.

6.3.

1. $|AB| = \sqrt{10}$, $\cos(ABC) = \frac{1}{\sqrt{5}}$
2. $|AB| = 1$, $\cos(ABC) = e^{-8}$

6.4. $K(x, y) = 1 + x_1y_1 + x_2y_2 + 3x_1x_2 \cdot 3y_1y_2 + 2x_1^2 \cdot 2y_1^2 + 4x_2^2 \cdot 4y_2^2$

6.5. $f(x_1, x_2) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$

6.6. Ядром является только функция в пункте 1.

6.7.

6.8.

7.1.

7.2.

7.3.

7.4.

7.5.

7.6.

7.7.

8.1.

8.2.

8.3.

8.4.

9.1. $I_X = 1 - 0.2^2 - 0.8^2 = 0.32$, $H(X) = -(0.2 \ln 0.2 + 0.8 \ln 0.8) \approx 0.5$
$I_Y = 1 - 0.2^2 - 0.3^2 - 0.5^2 = 0.62$, $H(Y) = -(0.2 \ln 0.2 + 0.3 \ln 0.3 + 0.5 \ln 0.5) \approx 1.03$

9.2.

1. $H(X) = \ln 3$, $I_X = 2/3$, спутанность равна 3.

2. $I_X = 1 - \frac{1}{k}$, $H(X) = \ln k$, спутанность равна k.

3. Если величина X равновероятно принимает k значений, то спутанность равна k. У равномерной на $[0; a]$ спутанность равна a. $H(X) = -\int_0^a \frac{1}{a} \cdot \ln \frac{1}{a} dx = \ln a$.

4. Обозначим $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, тогда $H(X) = -\int_{-\infty}^{+\infty} f(x) \ln(f(x)) dx$.

9.3.

9.4. $I = 2p(1-p)$, энтропия и индекс Джини максимальны при $p = 0.5$.

9.5.

$$I = 2 \cdot 0.7 \cdot 0.3$$

9.6.

$$I = 0.2 I_L + 0.8 I_R$$

9.7.

9.8.

9.9. Было: $I = 1 - \left(\frac{20}{70}\right)^2 - \left(\frac{40}{70}\right)^2 - \left(\frac{2}{70}\right)^2 - \left(\frac{8}{70}\right)^2 = \frac{708}{1225} \approx 0.58$,
$H = -\left(\frac{20}{70}\ln\frac{20}{70} + \frac{40}{70}\ln\frac{40}{70} + \frac{2}{70}\ln\frac{2}{70} + \frac{8}{70}\ln\frac{8}{70}\right) \approx 1.03$.
Стало: $I_L = 0$, $I_R = 1 - \left(\frac{20}{30}\right)^2 - \left(\frac{2}{30}\right)^2 - \left(\frac{8}{30}\right)^2 = 0.48$, $I = \frac{40}{70} \cdot 0 + \frac{30}{70} \cdot 0.48 \approx 0.21$,
$H_L = 0$, $H_R = -\left(\frac{20}{30}\ln\frac{20}{30} + \frac{2}{30}\ln\frac{2}{30} + \frac{8}{30}\ln\frac{8}{30}\right) \approx 0.8$, $H = \frac{40}{70} \cdot 0 + \frac{30}{70} \cdot 0.8 \approx 0.34$.

9.10.

9.11. $100 \cdot \left(\frac{99}{100}\right)^{100} \approx 100/e \approx 37$

9.12. Сначала делим по z, потом по x, так как индекс Джини в таком порядке падает сильнее.

9.13.

9.14. Нет, в силу выпуклости функций.

9.15. Все y_i одинаковые; поровну y_i двух типов; 1000 разных типов y_i, по одному наблюдению каждого типа.

9.16.

y_i	x_i	z_i
1	1	1
1	2	2
0	1	2
0	2	1

9.17.

9.18.

11 Источники мудрости