1 Линейная регрессия

1.1 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 $\hat{y} = H \cdot y$.

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

1.3 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?

2 Линейные классификаторы

2.1 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.

2.2 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).

2.3 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.

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

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

2.5 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.

2.6 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).

[1] деревня в Кадуйском районе Вологодской области

4. Найдите площадь под PR-кривой.
5. Как по-английски будет «бревно»?

2.7 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.

2.8 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.

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

2.9 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.

2.10 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.

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

2.12 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?

2.13 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!

2.14 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?

2.15 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?

2.16 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?

2.17 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?

3 Матрицы

3.1 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;

3.2 Объясните геометрический смысл QR, SVD и спектрального разложений.

4 Метод опорных векторов

4.1 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 $\sigma = 1$.
3. Найдите расстояние AB и косинус угла ABC в расширенном пространстве с помощью полиномиального ядра второй степени.

4.2 Переход из двумерного пространства в расширяющее задан функцией

$$f : (x_1, x_2) \to (1, x_1, x_2, 3x_1x2, 2x_1^2, 4x_2^2).$$

Найдите соответствующую ядерную функцию.

4.3 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?

4.4 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.

4.5 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.

4.6 Ядерная функция, скалярное произведение в расширяющем пространстве, имеет вид $K(\vec{a}, \vec{b}) = \exp(-|\vec{a} - \vec{b}|^2)$.

Имеются вектора $\vec{a} = (1, 1, 1)$ и $\vec{b} = (1, 2, 0)$.

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

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 σ, $k(v_1, v_2) = \exp(-\sigma|v_1 - v_2|^2)$.

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

4.8 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?

4.9 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 $\langle w, x\rangle = w_0 - 1$.

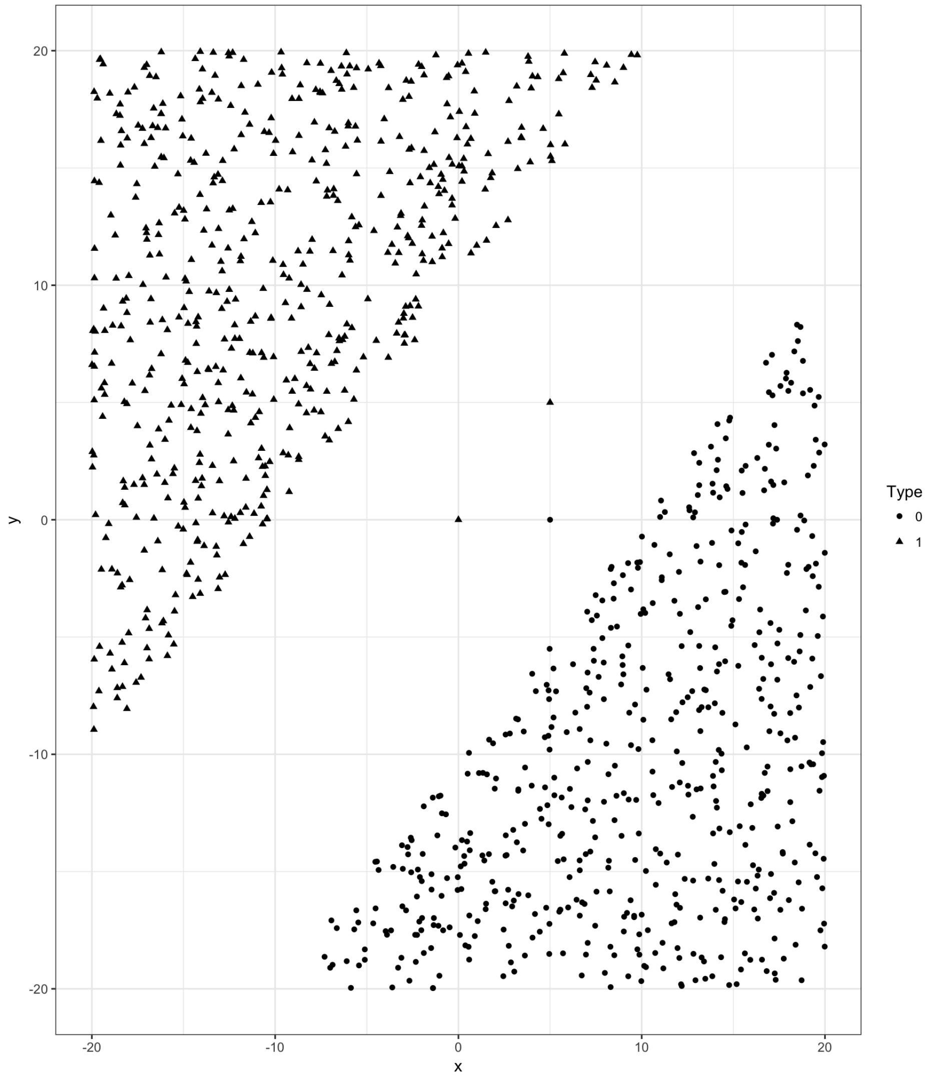
4.10 Заданы две прямые, l_0: $x^{(1)} + 3x^{(2)} = 9$ и l_1: $x^{(1)} + 3x^{(2)} = 13$. Найдите подходящий вектор w и число w_0 так, чтобы прямая l_0 записывалась как $\langle w, x\rangle = w_0 - 1$, а прямая l_1 как $\langle w, x\rangle = w_0 + 1$.

4.11 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 $\hat{y}$.

4.12 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.

4.13 юююю