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

2.4 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.

2.5 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).

2.6 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.

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

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

2.8 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.

2.9 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»?

2.10 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.

2.11 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?

2.12 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.

2.13 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.

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

2.15 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?

2.16 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!

2.17 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?

2.18 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?

2.19 юююю

2.20 юююю