

Тятя! Тятя! Наши сети притащили мертвеца!

эконом РАНХиГС
осень 2019

Задачи к посиделке 2

Задача 1

У Бандерлога три наблюдения¹, первое наблюдение — кит, остальные — муравьи. Киты кодируются $y_i = 1$, муравьи — $y_i = 0$. В качестве регрессоров Бандерлог берёт номера наблюдений $x_i = i$. 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?

Задача 2

Бандерлог чуть внимательнее присмотрелся к своему третьему наблюдению и понял, что это не кит, а бобёр. Теперь ему нужно решать задачу классификации на три класса. Он решил использовать для этого нейросеть с softmax-слоем на выходе. Предположим, что сетка обучилась и на двух новых наблюдениях, перед самым softmax-слоем она выплонула 1, 2, 5 и 2, 5, 1.

1. Чему равны вероятности получить кита, муравья и бобра для обеих ситуаций?
2. Пусть первым был кит, а вторым бобёр. Чему будет равна logloss-ошибка?

Задача 3

Иногда в функцию Softmax добавляют дополнительный параметр T , который называют температурой. Тогда она приобретает вид

$$f(z) = \frac{e^{\frac{z_i}{T}}}{\sum_{k=1}^K e^{\frac{z_k}{T}}}$$

Обычно это делается, когда с помощью нейросетки нужно сгенерировать какой-нибудь новый объект. Пусть у нас есть три класса. Наша нейросеть выдала на последнем слое числа 1, 2, 5.

¹Про другие приключения Бандерлога читай тут: https://github.com/bdemeshev/mlearn_pro/blob/master/mlearn_pro.pdf

1. Какое итоговое распределение вероятностей мы получим, если $T = 10$?
2. А если $T = 1$?
3. А если $T = 0.1$?
4. Какое распределение получится при $T \rightarrow 0$?
5. А при $T \rightarrow \infty$?
6. Предположим, что объектов на порядок больше. Например, это реплики, которые Алиса может сказать вам в ответ на какую-то фразу. Понятное дело, что вашей фразе будет релевантно какое-то подмножество ответов. Какое значение температуры сэмплирования T смогут сделать реплики Алисы непредсказуемыми? А какие сделают их однотипными?

Задача 4

Функция $f(t) = \frac{e^t}{1+e^t}$ называется сигмоидой².

1. Что происходит при $t \rightarrow +\infty$? А при $t \rightarrow -\infty$?
2. Как связаны между собой $f(t)$ и $f(-t)$?
3. Как связаны между собой $f'(t)$ и $f'(-t)$?
4. Как связаны между собой $f(t)$ и $f'(t)$?
5. Найдите $f(0)$, $f'(0)$ и $\ln f(0)$.
6. Найдите обратную функцию $f^{-1}(t)$
7. Как связаны между собой $\frac{d \ln f(t)}{dt}$ и $f(-t)$?
8. Постройте графики функций $f(t)$ и $f'(t)$.
9. Разложите $h(\beta_1, \beta_2) = \ln f(y_i(\beta_1 + \beta_2 x_i))$ в ряд Тейлора до второго порядка в окрестности точки $\beta_1 = 0, \beta_2 = 0$.

Задача 5

Маша Нестерова, хозяйка машин лёрнинга, собрала два наблюдения: $x_1 = 1, x_2 = 2, y_1 = 2, y_2 = 3$ и собирается обучить линейную регрессию $y = \beta \cdot x$. Маши очень хрупкая девушка, и ей не помешает помощь.

1. Получите теоретическую оценку методом наименьших квадратов.
2. Сделайте два шага градиентного спуска. В качестве стартовой точки используйте $\beta_0 = 0$. В качестве скорости обучения возьмите $\eta = 0.1$.
3. Сделайте два шага стохастического градиентного спуска. Пусть в SGD сначала попадает первое наблюдение, затем второе.
4. Если вы добрались до этого пункта, вы поняли градиентный спуск. Маша довольна. Начинаем заниматься тупой технической бессмыслицей. Сделайте два шага Momentum SGD. Возьмите $\alpha = 0.9, \eta = 0.1$
5. Сделайте два шага Momentum SGD с коррекцией Нестерова.
6. Сделайте два шага RMSprop. Возьмите $\alpha = 0.9, \eta = 0.1$

²В этом всё тоже замешан один Бандерлог.

7. Шоб ещё такого сделать? Придумал! Давайте сделаем два шага Adam! Возьмём $\beta_1 = \beta_2 = 0.9, \eta = 0.1$

Задача 6

Иногда в задаче классификации вместо меток $\{0, 1\}$ используют метки $\{-1, 1\}$. Как при таком изменении меток поменяется функция потерь? Выведите её из метода максимального правдоподобия. Как два полученных \logloss связаны между собой?

Задача 7

Всю свою жизнь Филипп был бездельником и сидел на шее у родителей. Но наконец ему улыбнулась удача и его взяли на работу аналитиком. Теперь он борется со всем плохим и за всё хорошее внутри большой рекомендательной системы.

Когда пользователь рекомендательной системы встречается в статьях матерную брань, он очень сильно расстраивается. Из-за этого одной из задач Филиппа является воспитание модели, которая могла бы дать бой нецензурщине.

Филипп оценил модель и получил на тестовой выборке следующие прогнозы вероятностей:

y_i	\hat{p}_i
1	0.99
1	0.8
0	0.77
1	0.55
0	0.55
0	0.4
1	0.3
0	0.2

В первой колонке стоит 1, если статья была матерной и стоит 0, если статья оказалась обычной (это реальность). Во второй колонке стоит вероятность того, что статья окажется матерной, предсказанная моделью. Филипп размышляет над тем какой бы для модели выбрать порог: 0.5, 0.7 или 0.9.

1. Найдите для каждого порога precision и recall. Какой порог вы бы порекомендовали поставить? Обоснуйте ответ.
2. Что такое гос-аус? Найдите значение этой метрики. Правда ли, что оно не зависит от выбранного порога?