

ДЗ №4

Даниил Дмитриев, 494

11 апреля 2017 г.

1 Задание 3.1

1. Как выглядит бинарный линейный классификатор? (Формула для отображения из множества объектов в множество классов.)

Он отображает x в $\text{sign} \langle w, x \rangle$, где w - матрица весов. (Считаем, что у x есть фиктивная единичная координата).

2. Что такое отступ алгоритма на объекте? Какие выводы можно сделать из знака отступа?

Отступ - $y_i \langle w, x_i \rangle$, он положителен тогда и только тогда, когда мы правильно предсказали класс объекта

3. Как классификаторы вида $a(x) = \text{sign}(\langle w, x \rangle - w_0)$ сводят к классификаторам вида $a(x) = \text{sign}(\langle w, x \rangle)$?

Добавляют фиктивную единичную координату к вектору признаков x .

4. Как выглядит запись функционала эмпирического риска через отступы? Какое значение он должен принимать для «наилучшего» алгоритма классификации?

$$Q = \sum_{i=1}^n I\{M_i \leq 0\}, \text{ для "наилучшего" алгоритма } L = 0$$

5. Если в функционале эмпирического риска (риск с пороговой функцией потерь) всюду написаны строгие неравенства ($M_i < 0$) можете ли вы сразу придумать параметр w для алгоритма классификации $a(x) = \text{sign}(\langle w, x \rangle)$, минимизирующий такой функционал?

w - нулевая матрица

6. Запишите функционал аппроксимированного эмпирического риска, если выбрана функция потерь $L(M)$.

$$Q = \sum_{i=1}^n L(M_i)$$

7. Что такое функция потерь, зачем она нужна? Как обычно выглядит ее график?

Функция потерь - функция, показывающая степень ошибки нашего алгоритма, её мы минимизируем. График зависит от того, строим мы его

на обучающей или на тестовой выборке, и от того, от каких осей мы его откладываем. На обучающей он обычно убывает и стремится к какому-то пределу, на тестовой он сначала убывает, а потом увеличивается из-за переобучения.

8. Приведите пример негладкой функции потерь.

$$L = \sum_{i=1}^n |\hat{y}_i - y_i|$$

9. Что такое регуляризация? Какие регуляризаторы вы знаете?

Регуляризация - штраф за возможное переобучение (например, большие веса признаков). Возможные регуляризаторы - L_2, L_1

10. Как связаны переобучение и обобщающая способность алгоритма? Как влияет регуляризация на обобщающую способность?

Чем сильнее алгоритм переобучился, тем хуже у него обобщающая способность. Регуляризация добавляется с целью улучшить обобщающую способность.

11. Как связаны острые минимумы функционала аппроксимированного эмпирического риска с проблемой переобучения?

Из-за них увеличивается риск переобучения

12. Что делает регуляризация с аппроксимированным риском как функции параметров алгоритма?

Добавляет ограничение на величину этих параметров, то есть в идеале рассматривает эту функцию только внутри какой-то области.

13. Для какого алгоритма классификации функционал аппроксимированного риска будет принимать большее значение на обучающей выборке: для построенного с регуляризацией или без нее? Почему?

Без регуляризации, так как без неё мы сможем лучше подстроиться под обучающие данные

14. Для какого алгоритма классификации функционал риска будет принимать большее значение на тестовой выборке: для построенного с опра-
дывающей себя регуляризацией или вообще без нее? Почему?

С регуляризацией, так как с ней у алгоритма улучшается обобщающая способность и мы не сильно подстраиваемся под обучающую выборку

15. Что представляют собой метрики качества Accuracy, Precision и Recall?

Accuracy - доля правильных ответов, Precision - отношение размера правильно угаданного класса 1 к всему предсказанному классу 1, Recall - отношение размера правильно угаданного класса 1 к всему настоящему классу 1

16. Что такое метрика качества AUC и ROC-кривая?

AUC - площадь под ROC-кривой, ROC-кривая - кривая, построенная по следующему принципу: берём список предсказаний, варьируем порог, начиная с которого мы относим к 1 классу, от 0 до 1 и наносим точки (precision,

recall) на график.

17. Как построить ROC-кривую (нужен алгоритм), если например, у вас есть правильные ответы к домашнему заданию про фамилии и ваши прогнозы?

Сортирую свои прогнозы, в цикле ставлю порог равным очередному прогнозу, с помощью правильных ответов узнаю precision, recall, наношу их на график.

2 Задача 3.3

Было на лекции. Добавляются $\xi_i \leq 0$ - допустимые ошибки. Сводится к линейной классификации с hinge loss и l2 регуляризацией, с наблюдением, что $\xi_i = (1 - M_i)_+$ где M_i - отступ.

3 Задача 3.4

$K(w_0, x) = x_1^2 + 2x_2^2 - 3 = \langle \psi(w_0), \psi(x) \rangle$ Попробуем подобрать $\psi(x)$

$$\psi(x) = (x_1^2, 2x_2^2, 1), \psi(w_0) = (1, 1, -3)$$

Получаем то, что и требовалось. Размерность пространства - 3.

4 Задание 3.6

См. номера 15, 16, 17 из задания 3.1