Антон Каразеев, 493

3. Теоретические задачи.
3.1 Знакомство с линейным классификатором

1. Как выглядит бинарный линейный классификатор?

   Есть два класса объектов A={-1, +1}. Отображение $f(x) : X \to A$ называется классификатором, отображающим объекты из множества X во множество классов A. Линейный классификатор выглядит следующим образом: $f(x) = sign(w^T x + w_0)$.

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

   В общем виде отступ $M(x_i) = y_i g(x_i)$, где y_i - метка i-того класса. Так как множество классов A={-1, +1}, то можно сделать вывод о том, что при правильном отнесении объекта к классу $M(x_i)$ положителен. В противном случае - отрицательный. Следовательно, неположительный отступ - ошибка классификатора.

3. Как классификаторы вида $a(x) = sign(< w, x > -w_0)$ сводят к классификаторам вида $a(x) = sign(< w, x >)$?

   К вектору x добавляют еще одну координату со значением -1, а к вектору w — w_0.

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

   $Q(X) = \sum_{x \in X} I\{M(x) < 0\}$

   Для "наилучшего"алгоритма классификации он должен принимать значение 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 $a(x) = sign(< w, x >)$, минимизирующий такой функционал?

   Положить $w = 0$.

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

   $Q(X) = \frac{1}{|X|} \sum_{x \in X} L(M(x))$

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

   Это неотрицательная функция, отражающая величину ошибки классификатора f на объекте x. График - монотонная функция.

8. Приведите пример негладкой функции потерь.

   $$L(x) = ReLU(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

9. Что такое регуляризация? Какие регуляризаторы вы знаете?

   Регуляризация штрафует за "усложнение"модели.

   $Q(X) = \frac{1}{|X|} \sum_{x \in X} L(M(x)) + \lambda Reg(w)$

   E.g. $Reg(w) = \sum_{i=1}^{|w|} ||w_i||_1$ — ℓ_1 регуляризация, $Reg(w) = \sum_{i=1}^{|w|} ||w_i||_2$ — ℓ_2 регуляризация.

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

    Если алгоритм переобучен, то его обобщающая способность падает — например, если на множестве X_1 значение $Q(X_1)$ близко к нулю (классификатор был обучен на X_1), а на X_2 значение $Q(X_2)$ велико, то алгоритм классификации переобучен.

    Одним из факторов переобучения является наличие в векторе весов w координаты с большим значением (по модулю). Именно для борьбы с этим и существуют регуляризаторы.

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

    Любое смещение из такого минимума приведет к значительному росту функции $Q(X)$. Что не позволит найти глобальный минимум — обобщающая способность классификатора будет невысокой.

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

    Штрафует за "большие"значения параметров или за такие значения, которые выходят за границы.

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

    С регуляризацией, так как дополнительно будут суммироваться значения весов (сумма неотрицательна).

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

    Может быть и так, и так. Потому что алгоритм может переобучаться и показывать плохие результаты на тестовой выборке, а с регуляризацией — сумма штрафов вместе со значениями функции потерь будет равняться "плохим результатам"в предыдущем случае.

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

    Accuracy $= \frac{TP+TN}{\text{Total number of objects}}$

    Precision $= \frac{TP}{TP+FP}$

    Recall $= \frac{TP}{\text{Number of positive objects}}$

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

    ROC-кривая — график зависимости $TPR(FPR)$, где $TPR = \frac{TP}{\text{Number of positive objects}}$ и $FPR = \frac{FP}{\text{Number of negative objects}}$.

    AUC — area under curve — площадь под графиком ROC-кривой.

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

    Сортируем объекты $x \in X$ по значению дискриминантной функции (e.g. $< w, x >$).

Начинаем строить ROC-кривую из точки (0,0).

Пробегаем по отсортированному массиву: если класс объекта равен -1, двигаемся вправо, иначе — вверх.