machine learning $hw_0 1$

Ренат Камартдинов

March 2024

1. Опишите задачу машинного обучения. Дайте определение объекту, целевой переменной, признакам, модели, функционалу ошибки

   (a) Машинное обучение решает задачу нахождение неявных зависимостей там, где обычный свод правил может не работать. Пример: прогнозирование цен на недвижность, считывание эмоций, рекомедации книг и т.д.

   (b) Объект - то для кого(чего) мы делаем прогнозы/предсказания.

   (c) Целевая переменная - то что мы предсказываем

   (d) Признак - какая-то характеристика нашего объекта.

   (e) Модель - то, что предсказывает, делает прогноз. Модель отображает множество объектов во множество целевых переменных.

   (f) Функционал ошибки - мера качество работы модели на выборке.

2. Чем отличается функция потерь от функционала ошибки?
   Функция потерь - мера корректности модели ответа для **1 объекта**, а функционал ошибки - мера качество работы модели **на выборке**.

3. Какие функции потерь используются при решении задачи регрессии?

   (a) $(a(x) - y)^2$

   (b) $\sqrt{(a(x) - y)^2}$

   (c) MAE(a, X) = $\frac{1}{l} \sum_{i=1}^{l} |a(x_i) - y_i|$

   (d) Функция потерь Хуберта

$$L_\delta(a, x) = \begin{cases} \frac{1}{2}(y - a)^2, |y - a| < \delta \\ \delta(|y - a| - \frac{1}{2}\delta), |y - a| \geq \delta \end{cases}$$

   (e) $(log(a(x) + 1) - log(y + 1))^2$

4. Запишите формулу для линейной модели регрессии a(x) = w_0+w_1*$x1$+...+...w_d*x_d, где x - это значение признака, w - вес этого признака

5. Чем отличаются функционалы MSE и MAE? В каких случаях лучше использовать MSE, а в каких MAE?

   (a) MSE - очень сильно реагирует на выбросы, а MAE - не дифференцируема

   (b) Когда нужно продифференцировать функционал ошибки используете MSE, а когда в выборке много выбросы MAE

6. Чем отличается MAE от MAPE? Что более понятно заказчику продукта?
   MAPE - это функционал ошибки, который хорошо использовать когда нужно делать прогноз не зависяший от масштаба. MAE более понятен для заказчика так ответом будет велична знакомая закачика.

7. Что такое коэффициент детерминации? Как интерпретировать его значения? $R^2(a, X) = 1 - \frac{\sum_{i=1}^{l}(a(x_i)-y_i)^2}{\sum_{i=1}^{l}(y_i-\hat{y})^2}$
   Чем коэффицент детерменации ближе к 1, тем лучше модель

8. Чем log-cosh лучше функции потерь Хубера? Опишите обе функции потерь?

   (a)

$$L_\delta(y, a) = \begin{cases} \frac{1}{2}(y-2)^2, |y-a| < \delta \\ \delta(|y-a| - \frac{1}{2}\delta), |y-a| \geq \delta \end{cases}$$

   - Huber Loss

   (b) L(y, a) = logcosh(a - y) - logcosh function

   Имеет вторую производную

9. Что такое градиент? Какое его свойство используется при минимизации функций?
   Градиентом функции многих переменных в данной точке называется вектор, координаты которого равны частным производным по соотвествуюшим аргументам, вычисленным в данной точке
   $\Delta f(x_1, ..., x_d) = (\frac{df}{dx_j})_{i=1}^{d}$
   Градиент является направление скорейшого роста.

10. Что такое градиентный спуск? Опишите процесс алгоритма. Градиентный спуск - алгортим поиска точки минимума в функции.
    Описание алгоритма $w^{(0)}$ - Начальное приближение векторов веса
    $\nabla_w Q(w)$ - Градиент Q по w

    (a) Инициализируем вектор весов случайнами значениями

(b) Идем с шагом до тех пор пока значения не начнут нас удовлетворять
$w^(k) = w^(k-1) - \eta_k \nabla - wQ(w^{(k-1)}$

Когда остановить градиентный спуст?

(a) Когда ошибка на тестовой выборке перестает уменьшаться

(b) При слишком малом изменении весов на последней итерации

(c) Не сильно поменялась на обучающей выборке

11. Почему не всегда можно использовать полный градиентный спуск? Какие способы оценивания градиента вы знаете? Почему в стохастическом градиентном спуске важно менять длину шага по мере итераций? Какие стратегии изменения шага вы знаете?

(a) Потому что нам придется на каждом шаге вычислять градиент всей суммы.

(b) Вместо градиента всей функции можно подставлять градиент 1 случайно взятого слагаемого

(c) Потому что при приблежении к точке минимума мы можем пропустить эту точка, поэтому выжно уменьшать длину шагу

(d) $\eta_k = \frac{1}{k}$

12. Что такое переобучение? Как можно отследить переобучение модели?
Переобучение модели - это когда функционал ошибки на обучающий выборке существенно лучше, чем на тестовой выборке. Отложенная выборка и кросс валидация.

13. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?
    Кросс-валидация - это способ тестирования модели. Мы делим нашу выборку на k - блоков. Потом обучаем нашу модель на k - 1 блоке. И тестируем на k блоке. Повторяем нашу процедуру k раз каждый раз выбирая новый блок для тестирования. На размер тестовой выборкы

14. Как построить итоговую модель после того, как по кросс-валидации подобраны оптимальные гиперпараметры?
    Сделать композицию моделей. Если это задача регресси можно взять среднее арифмитечиское, а если это задача классификации можно взять наиболее часто встречающиеся ответы модели

15. Что такое регуляризация? Для чего используется?
    Регулязиция - это некая функция от w умноженная на констатнту. Для запрета модели использование больших весов.

16. Опишите, как работают L1- и L2-регуляризаторы.

    (a) $\lambda \sum_{i=1}^{d} |w_i|$ - L1 - регулизатор
    (b) $\lambda \sum_{i=1}^{d} w_i^2$

    L1 - регулязация отбирает признаки, а L2 - отбирает признаки.

17. Почему L1-регуляризация отбирает признаки?
    Потому что она может уменьшить вес признака до нуля.

18. Почему плохо накладывать регуляризацию на свободный коэффициент?

Потому что свободный коэффицент задает порядки величин.