Школа анализа данных
Машинное обучение, часть 2
Домашнее задание №1

Решите предложенные задачи. Решения необходимо оформить в виде PDF документа. Каждая задача должна быть подробно обоснована, задачи без обоснования не засчитываются. Решения пишутся в свободной форме, однако так, чтобы проверяющие смогли разобраться. Если проверяющие не смогут разобраться в решении какой-нибудь задачи, то она автоматически не засчитывается. Дедлайн очников 9 октября 2018 09:00MSK, дедлайн заочников и филиалов +2 суток.

Задача 1 (0.5 балла) Нейронные сети.

Реализуйте следующую булеву функцию трех переменных с помощью нейронной сети:

$$f(x) = (\bar{x}_1 \vee x_3) \& (x_1 \vee x_2 \vee x_3).$$

Все нейроны должны использовать функцию активации $\sigma(x) = [x > 0]$.

Задача 2 (1.5 балла) Нейронные сети, back-prop.

Рассмотрим двуслойную полносвязную нейронную сеть, применяемую для задачи классификации. На вход нейронной сети подается вектор признаков x размерности n, полносвязный слой с матрицей весов W размерности $n \times d$ преобразует вектор x в скрытое представление h некоторой размерности d:

$$h = xW$$

Функции активации нет, еще один полносвязный слой с матрицей весов W' размерности $d \times m$ преобразует скрытое представление в вектор оценок a принадлежности к каждому классу. Чтобы получить из этих оценок вероятности, используется softmax. Например, вероятность того, что объект, описываемый вектором признаков x, относится к классу j согласно нейронной сети выглядит так:

$$p_j = \frac{\exp(a_j)}{\sum_{k=1}^{m} \exp(a_k)}$$

В качестве функции потерь используется cross-entropy loss:

$$\mathcal{L} = -\sum_{j=1}^{m} y_j \log p_j,$$

где y – one-hot encoding истинной метки объекта.

Итак, мы полностью описали проход по нейронной сети вперед: как по входному вектору x найти вероятности классов p_j и вычислить значение функции потерь, зная ответ y на рассматриваемом объекте. Опишите обратный проход по нейронной сети: выпишите формулы изменения матриц весов W и W' в стохастическом градиентном спуске для метода обратного распространения ошибки (backpropagation).

Задача 3. Нейронные сети, инициализация весов.

Рассмотрим полносвязный слой нейронной сети с матрицей весов W и свободным членом b, получающий на вход вектор x размерности n и вычисляющиющий скрытое представление размерности m

$$h = Wx + b.$$

Предложите, из какого невырожденного вероятностного распределения надо выбирать веса W и b, чтобы активации h имели нормальное распределение $N(0, \sigma^2)$, если

(a) **(1 балл)** Все признаки независимы и распределены по стандартному нормальному закону.

(b) **(2 балла)** Все признаки независимы и распределены равномерно от 0 до a.

Распределения W и b не обязаны совпадать, они могут быть из разных семейств.

Задача 4 (1.5 балла) Композиции алгоритмов, бустинг, AdaBoost.

Обозначим через $\tilde{w}^{(N)}$ нормированный вектор весов на N-й итерации алгоритма AdaBoost. Покажите, что взвешенная ошибка базового классификатора b_N относительно весов со следующего шага $\tilde{w}_i^{(N+1)}$ равна 1/2:

$$\sum_{i=1}^{\ell} \tilde{w}_i^{(N+1)} [b_N(x_i) \neq y_i] = \frac{1}{2}.$$

Задача 5 (2 балла) Градиентный бустинг.

1. Какой функции потерь будет соответствовать градиентный бустинг, который на каждой итерации настраивается на разность между вектором истинных меток и текущим вектором предсказанных меток?

2. Градиентный бустинг обучается на пяти объектах с функцией потерь для одного объекта

   $$\mathcal{L}(\tilde{y}, y) = (\tilde{y} - y)^4.$$

   На некоторой итерации полученная композиция дает ответ $(5, 10, 6, 3, 0)$. На какой вектор ответов будет настраиваться следующий базовый алгоритм, если истинный вектор ответов равен $(6, 8, 6, 4, 1)$?

3. Рассмотрим задачу бинарной классификации, $Y = \{0, 1\}$. Будем считать, что все алгоритмы из базового семейства $\mathcal{A}$ возвращают ответы из отрезка $[0, 1]$, которые можно интерпретировать как вероятности принадлежности объекта к классу 1. В качестве функции потерь возьмем отрицательный логарифм правдоподобия (negative log-likelihood):

   $$L(y, z) = -\big(y \log z + (1 - y) \log(1 - z)\big),$$

   где y — правильный ответ, а z — ответ алгоритма.

   Выпишите формулы для поиска базовых алгоритмов b_n и коэффициентов γ_n в градиентном бустинге.

Задача 6 (1.5 балла) Композиции, устойчивость к шуму.

1. Рассмотрим алгоритм AdaBoost — бустинг с экспоненциальной функцией потерь

   $$\mathcal{L}(M) = \exp(-M),$$

   где M — отступ объекта. Покажите, что алгоритм неустойчив к шуму, т.е. возможен неограниченный рост отношения весов шумовых объектов по отношению к весам пороговых объектов.

2. Покажите, что бустинг с логистической функцией потерь

$$\mathcal{L}(M) = \log(1 + \exp(-M))$$

устойчив к шуму в описанном выше смысле смысле.

Примечание. Пороговые объекты — это те, для которых значение отступа положительно и порядка нуля, то есть они лежат близко к границе между классами и в своем классе. Шумовые объекты лежат глубоко в чужом классе, на них отступ принимает большие отрицательные значения.