Машинное обучение, ФКН ВШЭ
SVM и логистическая регрессия

Задача 1. Роберта утверждает, что открыла новую дифференциируемую верхнюю границу для пороговой функции потерь,

$$\tilde{L}(M_i) = \frac{9}{10} - \frac{1}{\pi} \cdot \arctan(M_i),$$

где $M_i = y_i \cdot \langle w, x_i \rangle$. Права ли Роберта?

Задача 2. Позволяет ли предсказывать корректные вероятности экспоненциальная функция потерь $L(y, z) = \exp(-yz)$?

Задача 3. Рассмотрим постановку оптимизационной задачи метода опорных векторов для линейно разделимой выборки:

$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min_{w,b}, \\ y_i(\langle w, x \rangle + b) \geq 1, \quad i = \overline{1, \ell}, \end{cases}$$

а также её видоизменёный вариант для некоторого значения $t > 0$:

$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min_{w,b}, \\ y_i(\langle w, x \rangle + b) \geq t, \quad i = \overline{1, \ell}. \end{cases}$$

Покажите, что разделяющие гиперплоскости, получающиеся в результате решения каждой из этих задач, совпадают.

Задача 4. Ответьте на следующие вопросы:

1. Почему в общем случае распределение $p(y|x)$ для некоторого объекта $x \in \mathbb{X}$ отличается от вырожденного ($p(y|x) \in \{0, 1\}$)?

2. Почему логистическая регрессия позволяет предсказывать корректные вероятности принадлежности объекта классам?

3. Рассмотрим оптимизационную задачу из варианта SVM для линейно разделимых выборок. Всегда ли в обучающей выборке существует объект x_i, для которого выполнено $y_i(\langle w, x_i \rangle + b) = 1$? Почему?

4. С какой целью в постановке оптимизационной задачи SVM для линейно неразделимых выборок вводятся переменные ξ_i, $i = \overline{1, \ell}$?

Задача 5. Рассмотрим целевую функцию логистической регрессии

$$Q(w) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log(1 + \exp(-y_i \langle w, x_i \rangle)),$$

1. Найдите градиент ∇Q_w и упростите итоговое выражение таким образом, чтобы в нём участвовала сигмоидная функция

$$\sigma(z) = \frac{1}{1 + \exp(-z)}.$$

   При решении данной задачи вам может понадобиться следующий факт (убедитесь, что он действительно выполняется):

$$\sigma'(z) = \sigma(z)(1 - \sigma(z)).$$

2. Выпишите, как будет выглядеть шаг градиентного спуска.

3. Найдите вторую производную целевой функции по w.

4. Выпишите квадратичную аппроксимацию для $Q(w)$ в окрестности $w = 0$. Для этого разложите функцию потерь в ряд Тейлора до второго члена в окрестности точки $w = 0$. С какой задачей совпадает задача минимизации квадратичной аппроксимации?