Машинное обучение
Теоретическое домашнее задание №4

Задача 1. Позволяет ли предсказывать корректные вероятности экспоненциальная функция потерь $L(y, z) = \exp(-yz)$?

Задача 2. 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 3. Пусть мы решили двойственную задачу SVM и получили оптимальные значения $(\lambda_1, \dots, \lambda_\ell)$, где $\lambda_5 = C/3$, $\lambda_2 = 0$. Выразите оптимальное значение порога b для прямой задачи через найденное решение $(\lambda_1, \dots, \lambda_\ell)$ двойственной задачи.

Задача 4. Вычислите градиент $\frac{\partial}{\partial w} L(x, y; w)$ логистической функции потерь для случая линейного классификатора

$$L(x, y; w) = \log(1 + \exp(-y \langle w, x\rangle))$$

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):

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

Задача 5. 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 hard-margin SVM. 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 soft-margin SVM вводятся переменные ξ_i, $i = \overline{1, \ell}$?