

Домашнее задание по материалу 6-го семинара.
ММП, осень 2012–2013
12 октября

Задача 1. Пусть линейный классификатор $a(\mathbf{x})$ в задаче с двумя классами, где $\mathbf{x} \in \mathbb{R}^n$, имеет вид

$$a(x) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + w_0).$$

- Найдите нормаль к разделяющей поверхности классификатора.
- Найдите расстояние от центра координат до разделяющей поверхности.
- Найдите расстояние от объекта \mathbf{x}_0 до разделяющей поверхности.

Выпуклой оболочкой множества $X = \{x_1, \dots, x_m\} \in \mathbb{R}^n$ называется множество

$$\text{conv}(X) = \{x : x = \sum_{i=1}^m \alpha_i x_i; \sum_i \alpha_i = 1, \alpha_i \geq 0, i = 1, \dots, m\}.$$

Задача 2. Дано два множества точек $A, B \subset \mathbb{R}^n$. Докажите, что

- Если *выпуклые оболочки* множеств пересекаются ($\text{conv}(A) \cap \text{conv}(B) \neq \emptyset$), то множества A и B линейно не разделимы.
- Если множества A и B линейно разделимы, то их выпуклые оболочки не пересекаются.

Задача 3. Рассмотрим логистическую регрессию в одномерной задаче $\mathbb{X} = \mathbb{R}$ с двумя классами $\mathbb{Y} = \{0, 1\}$:

$$P(y = 1|x, \mathbf{w}) = \sigma(w_1 x + w_0),$$

где $\sigma(x) = \frac{1}{1+e^{-x}}$ — логистическая сигмоида.

На рисунке 1 приведены две разных функции апостериорных вероятностей $P(y = 1|x, \mathbf{w})$ принадлежности к классу 1, получающихся при различных параметрах \mathbf{w} .

- Для каждой из апостериорных вероятностей укажите число ошибок, допускаемых на объектах, приведенных на том же рисунке.
- Одна из приведенных апостериорных вероятностей соответствует вектору \mathbf{w} , полученному методом максимизации правдоподобия на объектах, указанных на рисунке. Которая из них?
- Повлияет ли на решение пункта б) добавление *регуляризатора* $w_1^2/2$ к логарифму функции правдоподобия?

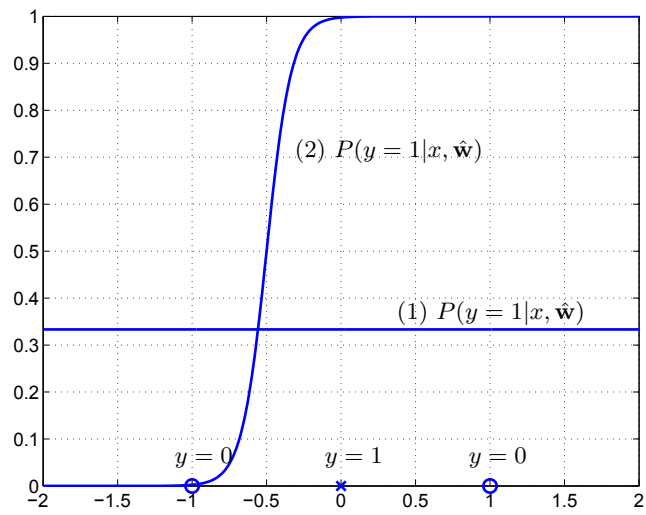


Рис. 1: Обучающие объекты.