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

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

$$a(x) = \operatorname{sgn}\big(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + w_0\big).$$

а) Найдите нормаль к разделяющей поверхности классификатора.
б) Найдите расстояние от центра координат до разделяющей поверхности.
в) Найдите расстояние от объекта $\boldsymbol{x_0}$ до разделяющей поверхности.

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

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

Задача 2. Дано два множества точек $A, B \subset \mathbb{R}^n$. Докажите, что
а) Если *выпуклые оболочки* множеств пересекаются ($conv(A) \cap conv(B) \neq \emptyset$), то множества A и B линейно не разделимы.
б) 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 $\mathbb{Y} = \{0, 1\}$:

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

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

На рисунке 1 приведены две разных функции апостериорных вероятностей $\mathsf{P}(y = 1 | x, \boldsymbol{w})$ принадлежности к классу 1, получающихся при различных параметрах $\boldsymbol{w}$.
а) Для каждой из апостериорных вероятностей укажите число ошибок, допускаемых на объектах, приведенных на том же рисунке.
б) Одна из приведенных апостериорных вероятностей соответствует вектору $\boldsymbol{w}$, полученному методом максимизации правдоподобия на объектах, указанных на рисунке. Которая из них?
в) Повлияет ли на решение пункта б) добавление *регуляризатора* $w_1^2/2$ к логарифму функции правдоподобия?

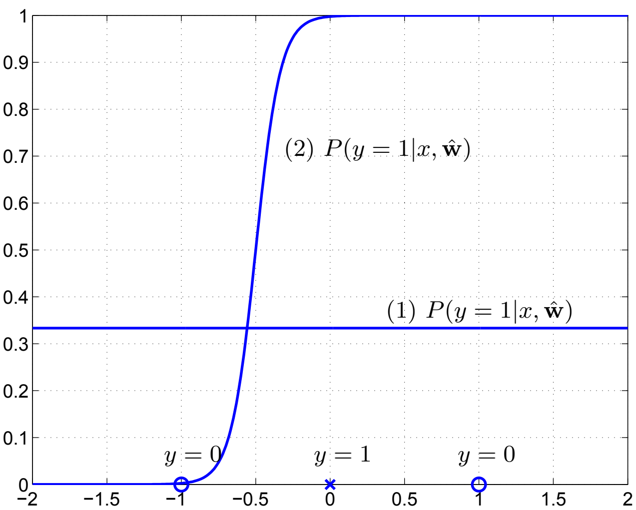

Рис. 1: Обучающие объекты.