

# Машинное обучение, ФКН ВШЭ

## Теоретическое домашнее задание №7

**Задача 1.** Покажите, что параметр  $\nu$ , используемый в постановке задачи одноклассового SVM, является верхней оценкой на долю аномалий на обучающей выборке, т.е. что  $\frac{1}{\ell} \sum_{i=1}^{\ell} [\langle w, x_i \rangle < \rho] \leq \nu$ , где  $w$  — оптимальный вектор весов.

**Задача 2.** В факторизационных машинах предсказание для объекта  $x \in \mathbb{R}^d$  делается по формуле

$$a(x) = w_0 + \sum_{j=1}^d w_j x_j + \sum_{j=1}^d \sum_{k=j+1}^d x_j x_k \langle v_j, v_k \rangle,$$

где  $w_j \in \mathbb{R}, v_j \in \mathbb{R}^r, j = \overline{0, d}$ . Вычисление предсказания по этой формуле требует  $O(rd^2)$  операций. Покажите, что это же предсказание может быть сделано за  $O(rd)$  операций.

**Задача 3.** Допустим, мы решили задать правдоподобие выборки через нормальное распределение для размеченных объектов и через смесь гауссиан для неразмеченных:

$$\begin{aligned} \sum_{i=1}^{\ell} \log \pi_{y_i} p(x_i | y_i, \theta) \\ + \sum_{i=\ell+1}^{\ell+k} \log \sum_{y \in \mathbb{Y}} \pi_y p(x_i | y, \theta) \rightarrow \max_{\theta}, \end{aligned}$$

где  $p(x | y, \theta) = \mathcal{N}(x | \mu_y, \Sigma_y)$ . Введите скрытые переменные и запишите формулы для Е- и М-шагов ЕМ-алгоритма.

**Задача 4.** Рассмотрим нормированный лапласиан — разновидность лапласиана, часто возникающую в литературе:

$$L_n = I - D^{-1/2} W D^{-1/2}.$$

Покажите, что для любого вектора  $f \in \mathbb{R}^{\ell}$  выполнено

$$f^T L_n f = \frac{1}{2} \sum_{i=1}^{\ell} w_{ij} \left( \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2.$$