Домашнее задание по материалу 10-го семинара. ММП, осень 2012–2013 11 декабря

1. Докажите, что если мы добавили в матрицу признаков A константный столбец, то для ответов на обучающей выборке $\hat{Y} = A\boldsymbol{\alpha}^*$ настроенного с помощью МНК алгоритма будет выполнено
$$\sum_{i=1}^{\ell}(Y_i - \hat{Y}_i) = 0.$$

2. Докажите, что вне зависимости от добавления константного признака, будет выполнено следующее равенство:
$$\sum_{i=1}^{\ell}\hat{Y}_i(Y_i - \hat{Y}_i) = 0.$$

3. Предположим, что целевая переменная (ответ) y выражается в виде
$$y = f(\boldsymbol{x}, \boldsymbol{\alpha}) + \varepsilon = \alpha_0 + \sum_{i=1}^{n}\alpha_i \boldsymbol{x}_i + \varepsilon, \tag{1}$$
где $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ — добавочный гауссовский шум. Таким образом, мы описали свойство распределения $\mathbb{P}$ на парах $(\boldsymbol{x}, y)$, из которого мы получаем объекты обучающей выборки.

   а) Зафиксируем квадратичную функцию потерь $\ell(y_1, y_2) = (y_1 - y_2)^2$. Докажите, что в предположениях (1) мат. ожидание квадратичных потерь $\mathbb{E}_{\boldsymbol{x},y}\ell\big(y, a(\boldsymbol{x})\big)$ минимизируется алгоритмом $a(\boldsymbol{x}) = \alpha_0 + \sum_{i=1}^{n}\boldsymbol{x}_i$.

   б) Пусть у нас есть обучающая выборка $X^\ell = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{\ell}$ из распределения (1). Докажите, что в этом случае метод максимума правдоподобия для поиска коэффициентов $(\alpha_0, \ldots, \alpha_n)$ эквивалентен методу наименьших квадратов.