# Среднее значение и стандартное отклонение ошибки

Задана выборка $\mathfrak{D} = \{\mathbf{x}_i, y_i\}$. Ее элементы проиндексированы:

$$i \in \mathcal{I} = \{1, \ldots, m\}.$$

Разобьем выборку равномерно случайно, на две равномощные подвыборки, обучение и контроль, $K$ раз:

$$\mathcal{I} \longrightarrow \mathcal{L}_k \bigsqcup \mathcal{C}_k, \qquad k \in \{1, \ldots, K\}.$$

Задана модель $f(\mathbf{w}, \mathbf{w})$ и функция ошибки $S(\mathbf{w}|\mathfrak{D})$. Параметры модели оптимизированы на обучении $\mathfrak{D}_{\mathcal{L}}$ как

$$\hat{\mathbf{w}} = \arg\min S(\mathbf{w}|f, \mathfrak{D}_{\mathcal{L}}).$$

Для каждого из $K$ разбиений вычисляем ошибку на обучении и на контроле. Получаем два набора ошибок:

$$\{S_k(\hat{\mathbf{w}}_k|f, \mathfrak{D}_{\mathcal{L}k})\}, \qquad \{S_k(\hat{\mathbf{w}}_k|f, \mathfrak{D}_{\mathcal{C}k})\}, \qquad k \in \{1, \ldots, K\}.$$

# Зависимость среднего значения ошибки от объема выборки

Для двух наборов ошибок вычислим среднее значение и поправленное стандартное отклонение:

$$\bar{S} = \frac{1}{K}\sum_{k=1}^{K} S_k, \qquad \sigma = \frac{1}{K-1}\sqrt{\sum_{k=1}^{K}(\bar{S} - S_k)^2}.$$

Повторим процедуру на ограниченном объеме выборки, например:

$$m = \overline{1, M},$$

где $M$ — наибольший объем доступной выборки.
Построим график зависимости ошибки и стандартного отклонения от объема выборки.
Домашнее задание 3 находится по адресу http://bit.ly/16UIlQH