

Задачи к семинару 10

Задача 1. Решить систему уравнений, используя метод регуляризации Тихонова $((X + \lambda I)\beta = y)$. Для подбора параметра регуляризации λ постройте график зависимости функционала Тихонова $\Omega = \|X\beta - y\|^2 + \lambda \|\beta\|^2$ от него и найдите на графике оптимальное значение. Внимание: график рекомендуется строить в логарифмических координатах.

$$\begin{cases} x + 7y & = 5 \\ \sqrt{2}x + \sqrt{98}y & = \sqrt{50} \end{cases}$$

Задача 2. Сгенерируйте выборку x из 100 равномерно распределённых случайных чисел ($x_i \in [0; 1]$). На её основе постройте выборки y и z следующим образом: $y_i = x_i$, $z_i = 3 + 2x_i + 5y_i + N(0; 1)$ ($N(0; 1)$ - случайная величина, подчиняющаяся стандартному нормальному распределению). Найдите параметры регрессии с помощью регуляризации Тихонова (Ridge regression) $\beta = (X^T X + \lambda I)^{-1} X^T z$. Попробуйте следующие значения параметра λ : 10^{-14} , 10^{-13} , 10^{-10} , 10^{-6} , 1, 1000.

Задача 3. Сгенерируйте набор точек в четырёхмерном пространстве (y, x_1, x_2, x_3) , используя следующий код на GNU Octave (в матрице xv i -й каждый столбец соответствует переменной x_i):

```
xv = nan(npoints, 3);
xv(:, 1) = 5*rand(npoints, 1);
xv(:, 2) = 0.7*(5*rand(npoints, 1)) + -0.3*xv(:, 1);
xv(:, 3) = 0.1*(5*rand(npoints, 1)) + 0.9*xv(:, 1);
yv = 3 + 4*xv(:, 1) + 5*xv(:, 2) + 6*xv(:, 3) + 8*randn(size(xv(:, 1))));
```

Выполните с ними следующие действия:

- Построить набор точечных графиков, на осях абсцисс и ординат которых находятся значения x_i и x_j ($i \neq j$). Для размещения нескольких графиков внутри одного окна можно пользоваться функцией `subplot`.
- Найти коэффициенты многомерной линейной регрессии (обычный МНК).
- Стандартизовать (отнормировать) точки и аппроксимировать их методом гребневой регрессии (Ridge regression). Построить зависимости стандартизованных коэффициентов регрессии и соответствующих им значений VIF (Variance Inflation Factor, фактор инфляции дисперсии) от λ (параметра регуляризации). Рекомендуемый диапазон значений $\lambda = 0 \div 1$

Стандартизация данных осуществляется по формулам:

$$y_i^* = \frac{y - \bar{y}}{\sqrt{\sum_j (y_j - \bar{y})^2}} = \frac{y - \bar{y}}{y_{\text{norm}}}; \quad x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_j (x_{ij} - \bar{x}_j)^2}} = \frac{x_{ij} - \bar{x}_j}{x_{j, \text{norm}}}$$

Значения VIF находятся на главной диагонали следующей матрицы (при условии стандартизации данных):

$$(X^T X + \lambda I)^{-1} (X^T X) (X^T X + \lambda I)^{-1}$$