Рассмотрим задачу безусловной оптимизации

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \gamma\|x\|_1 \to \min$$

При обучении Lasso регрессии мы придем к данной задаче оптимизации. Матрица A будет являться матрицей объектов-признаков, вектор x будет являться вектором весов линейной модели. Вектор b будет являться вектором значений целевой переменной.

Функция $f(x)$ не является дифференцируемой на всем $\mathbb{R}^d$, поэтому для решения поставленной задачи оптимизации нельзя применять стандартные алгоритмы.

Если сделать замену $z = Ax - b$, то функцию $f(x)$ можно будет заменить на функцию $f(x, z) = \frac{1}{2}\|z\|^2 + \gamma\|x\|_1$ при условии $z = Ax - b$.
Получим задачу условной оптимизации:

$$\begin{aligned} & f(x, z) = \frac{1}{2}\|z\|^2 + \gamma\|x\|_1 \to \min \\ & z = Ax - b \end{aligned}$$

Получение двойственной задачи оптимизации

Задача условной оптимизации регулярна, так как она выпукла и отсутствуют условия типа неравенств (применяем достаточное условие Слейтера при $m = 0$).

Для решения данной задачи попробуем составить двойственную задачу. Для этого запишем функцию Лагранжа:

$$\mathcal{L}(x, z, \mu) = \frac{1}{2}\|z\|^2 + \gamma\|x\|_1 + \mu^T(Ax - b - z)$$

Для сведения к двойственной задаче нужно проминимизировать $\mathcal{L}(x, z, \mu)$ по переменным $x \in \mathbb{R}^d$, $z \in \mathbb{R}^l$ для всех значений $\mu \in \mathbb{R}^l$. Функция Лагранжа не является дифференцируемой по z, поэтому не получится воспользоваться необходимым условием безусловного локального минимума.
Попробуем записать функцию Лагранжа в ином виде:

$$\mathcal{L}(x, z, \mu) = \frac{1}{2}\sum_i z_i^2 + \gamma\sum_j |x_i| + \sum_j (A^T\mu)_j x_j - \sum_i \mu_i z_i - \mu^T b$$

$$\mathcal{L}(x, z, \mu) = \sum_i \left(\frac{1}{2}z_i^2 - \mu_i z_i\right) + \sum_j \left(\gamma|x_i| - (A^T\mu)_j x_j\right) - \mu^T b$$

Мы смогли разбить функцию Лагранжа на $d + l$ независимых слагаемых. Поэтому минимизацию $\mathcal{L}$ можно свести к независимой минимизации всех слагаемых. Во все слагаемые входят μ_i, но при минимизации по x, z все μ_i --- зафиксированные параметры, через которые мы хотим выразить x и z.

Для начала проминимизируем слагаемые, отвечающие z_i:

$$\frac{1}{2}z_i^2 - \mu_i z_i \to \min_{z_i} \Rightarrow z_i = \mu_i$$

Тогда

$$\inf\left(\frac{1}{2}z_i^2 - \mu_i z_i\right) = -\frac{1}{2}\mu_i^2$$

Тогда

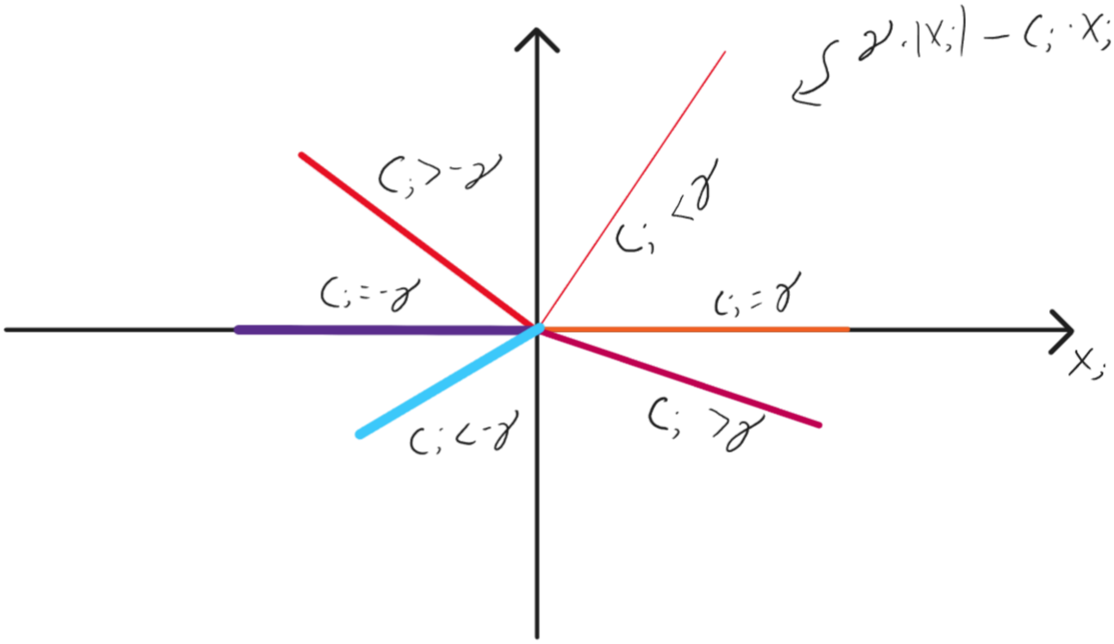
$$\inf\left(\sum_i \frac{1}{2}z_i^2 - \mu_i z_i\right) = -\frac{1}{2}\|\mu\|^2$$

Теперь проминимизируем слагаемые, отвечающие x_j. Данные слагаемые недифференцируемы, поэтому для минимизации придется использовать "метод пристального взгляда". Приведем слагаемые к более удобному виду и нарисуем графики получившихся одномерных функций при различных значениях μ:

$$h_j(x_j) = \gamma|x_j| - (A^T\mu)_j x_j$$

Для удобства обозначим $c_j := (A^T\mu)_j$. Тогда h_j примет следующий вид:

$$h_j(x_j) = \gamma|x_j| - c_j x_j = \begin{cases}(\gamma - c_j)x_j & , \quad x_j \geq 0 \\ (-\gamma - c_j)x_j & , \quad x_j < 0\end{cases}$$



Теперь проминимизируем функцию $h_j(x_j)$. Для этого проминимизируем по $x_j \geq 0$ и $x_j < 0$ и возьмем наименьшее значение

$$\inf_{x_j \in \mathbb{R}} h_j(x_j) = \min\{\inf_{x_j \geq 0} h_j(x_j), \inf_{x_j < 0} h_j(x_j)\}$$

$$\inf_{x_j \geq 0} h_j(x_j) = \begin{cases} 0, & \gamma - c_j > 0 \\ 0, & \gamma - c_j = 0 \\ -\infty, & \gamma - c_j < 0 \end{cases}$$

$$\inf_{x_j < 0} h_j(x_j) = \begin{cases} -\infty, & -\gamma - c_j > 0 \\ 0, & -\gamma - c_j = 0 \\ 0, & -\gamma - c_j < 0 \end{cases}$$

Ситуация $|c_j| < \gamma$ приводит к отбору признаков, так как в данном случае $\inf h_j(x_j)$ достигается только в точке $x_j = 0$

Таким образом, получим

$$\inf_{x_j} h_j(x_j) = \begin{cases} 0, & c_j \leq \gamma \ \wedge \ c_j \geq -\gamma \\ -\infty, & \text{иначе} \end{cases}$$

$$\inf_{x_j} h_j(x_j) = \begin{cases} 0, & |c_j| < \gamma \\ -\infty, & |c_j| \geq \gamma \end{cases}$$

$$\inf_x \sum_j \left(\gamma|x_i| - (A^T\mu)_j x_j\right) = \sum_j \inf_{x_j} h_j(x_j) = \begin{cases} 0, & \forall j \Rightarrow |c_j| < \gamma \\ -\infty, & \exists j: \quad |c_j| \geq \gamma \end{cases}$$

Если вспомнить, что $c_j = A^T\mu_j$, то мы получим выражение для $\inf \mathcal{L}(x, z, \mu)$:

$$g(\mu) = \inf_{x,z} \mathcal{L}(x, z, \mu) = \begin{cases} -\infty, & \exists j: \ |(A^T\mu)_j| \geq \gamma \\ -\frac{1}{2}\|\mu\|^2, & \forall j \Rightarrow -\gamma \leq (A^T\mu)_j \leq \gamma \end{cases}$$

Тогда получим следующую двойственную задачу:

$$\begin{aligned} & g(\mu) = -\frac{1}{2}\|\mu\|^2 \to \max \\ & (A^T\mu)_j \leq \gamma \ \ \forall j \\ & -(A^T\mu)_j \leq \gamma \ \ \forall j \end{aligned}$$

Получили задачу условной максимизации с линейными ограничениями и дифференцируемой целевой функцией!
Причем данная задача относится к классу задач квадратичного программирования. Для данного класса задач реализованы эффективные численные методы оптимизации, поэтому далее мы легко находим μ^* --- точку глобального максимума двойственной задачи.

После нахождения μ^* можно получить номера координат, которые обратятся в 0 из-за условия $|(A^T\mu^*)_j| = |c_j| < \gamma$. Далее остается найти значения ненулевых координат, для этого воспользуемся равенством $z = \mu^*$ и условием $Ax - b = z$. Отсюда получим, что x --- решение системы $Ax = \mu^* + b$. Перед запуском процедуры вычисления исключаем нулевые координаты.

В общем случае уравнение $Ax = \mu^* + b$ может не иметь решений, но так как μ^* --- решение двойственной задачи, то оно гарантирует выполнение равенства $Ax - b = z$

для оптимальной пары x, z. Поэтому данная система имеет решение, только его нужно аккуратно найти.