Домашнее задание по курсу «Методы оптимизации». ФИВТ. Тренин С.А. Осень 2016.
Срок сдачи 07.12.2016 23.59.

Отчет по заданию должен включать в себя:

- Код алгоритмов: (а) генерации входных данных, (б) градиентного спуска и выбора шага, (в) метода центрального пути, (г) демпфированного метода ньютона (для метода центрального пути), (д) вызов методов решения задач минимизации нормы вектора невязок (если для этого используется решение задачи ЛП, то реализовывать соответствующий решатель нет необходимости).
- Перечисленные в задании рисунки и поясняющий текст.
- Словесное (или на псевдокоде) описание всех алгоритмов, которые взяты из библиотек. Стандартные линейно-алгебраические алгоритмы (обращение или умножение матриц) описывать и реализовывать самостоятельно не следует.

1. (0.8 балла) Для идентификации параметров линейного классификатора методом логистической регрессии требуется найти минимум функции риска (http://www.machinelearning.ru/wiki/index.php?title=Логистическая_регрессия):

$$\mathrm{Q}(\omega) = \sum_{i=1}^{m} \ln(1 + \exp(-y_i \cdot < x_i, \omega >)$$

Для любого $\omega \in R^{n+1}$, где $x_i \in R^{n+1}$, $y_i \in \{-1, +1\}$, первая координата всех векторов x равна -1 ($x_i^0 = -1$). Остальные координаты и конкретные значения y требуется получить следующим образом:

- Смоделировать данные как точки двух классов $(-1, +1)$ на плоскости (n = 2) (разделимые прямой и не разделимые прямой).
- Тоже, но точки в трехмерном пространстве (n = 3). (Разделимые и не разделимые гиперплоскостью соответственно.)

Задание:

a) Реализовать алгоритм градиентного спуска средствами любого удобного для этих целей языка программирования, описать его для данной задачи и протестировать для различных начальных приближений. Отобразить на рисунках сгенерированное множество точек, гиперплоскость начального приближения и результирующую разделяющую гиперплоскость.
b) Значение градиента требуется посчитать аналитически (при выводе аналитических формул удобно использовать обозначение $\sigma(M) = \frac{1}{1+e^{-M}}$).
c) Построить графики зависимости количества итераций алгоритма от требуемой точности N(lg(1/ε)). При начальном одном и том же начальном приближении. Сравнить различные стратегии выбора шагов.
d) Построить разделяющую поверхность с помощью линейного дискриминантного анализа. При генерации данных используйте гипотезу линейного дискриминантного анализа (точки каждого класса выбраны из нормального распределения со своим математическим ожиданием и одинаковой дисперсией). (https://en.wikipedia.org/wiki/Linear_discriminant_analysis#LDA_for_two_classes) Сравнить результаты. Проведите оценку методом скользящего контроля.

2. (0.8 балла) Пусть дана задача ЛП в канонической форме.

$$\max c^T x$$
$$Ax + x_s = b$$
$$x, x_s \geq 0,$$
$$x \in \mathbb{R}^n, x_s, b \in \mathbb{R}^m, A - m \times \text{n матрица}$$
$$y \in \mathbb{R}^m, y_s \in \mathbb{R}^n - \text{переменные двойственной задачи}$$

Методом центрального пути называется следующий алгоритм. (Linear Programming Robert J. Vanderbei, стр. 277)

$$X \triangleq diag(x), X_s \triangleq diag(x_s), Y \triangleq diag(y), Y_s \triangleq diag(y_s),$$

$$\mathrm{F}(x, x_s, y, y_s, \mu) \triangleq \begin{bmatrix} Ax + x_s - b \\ A^T y - y_s - c \\ XY_s\mathbf{1} - \mu\mathbf{1} \\ YX_s\mathbf{1} - \mu\mathbf{1} \end{bmatrix}$$

$$\mu, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 > 0$$
$$(x, x_s, y, y_s) = \underset{x, x_s, y, y_s > 0}{solve}(\mathrm{F}(x, x_s, y, y_s, \mu_0) = 0)$$
$$repeat$$
$$\quad k = k + 1$$
$$\quad x^{prev} = x \quad , x_s^{prev} = x_s \ , y^{prev} = y \quad , y_s^{prev} = y_s$$
$$\quad (x, x_s, y, y_s) = \underset{x, x_s, y, y_s > 0}{solve}(\mathrm{F}(x, x_s, y, y_s, \mu \quad) = 0)$$
$$\quad \mu = 0.1\mu$$
$$until \left\|x^{prev} - x \quad \right\| \leq \epsilon_1 \& \left\|x_s^{prev} - x_s \ \right\| \leq \epsilon_2 \ \& \left\|y^{prev} - y \quad \right\| \leq \epsilon_3 \&$$
$$\left\|y_s^{prev} - y_s \ \right\| \leq \epsilon_4 \ or \ k > k\_max$$

Задание:

a) Показать, что, если алгоритм сходится, то он сходится к решения задачи ЛП $(x^k, x_s^k, y^k, y_s^k) \rightarrow (x^*, x_s^*, y^*, y_s^*)$, при $\mu \rightarrow 0$.

b) Реализовать алгоритм демпфированного метода Ньютона (варьируемая длина шага) для задачи $solve_{x, x_s, y, y_s > 0}(\mathrm{F}(x, x_s, y, y_s, \mu) = 0)$. Как выбрать демпфирующий параметр метода, для того чтобы гарантировать, что полученные решения будут неотрицательными?

c) Показать экспериментально, что с помощью построенного алгоритма можно решить задачи ЛП. Привести в отчете все входные данные, которые использовались для тестирования. Выберите ограниченную задачу оптимизации с двумя основными переменными (n=2). Решите эту задачу графическим или симплекс методом. Изобразите на плоскости (x_1, x_2) допустимое множество и решение (x_1^*, x_2^*) задачи. Отметьте на том же рисунки последовательность точек (x_1^k, x_2^k) и соответствующие значения μ.

d) Сформулируйте условие выхода, которое позволяет детектировать неограниченность задачи. Покажите экспериментально, что условие работает.

3. (0.4 балла) Пусть физический закон описывается зависимостью некоторого измеряемого значения y(x, a) от времени и координаты x при параметрах a:

$$y(t, a) = a_2 \sin(t) + a_1 t + a_0;$$

Дан набор координат t размера m. (t = {0 10/m 20/m ... (m-1)10/m} значения распределены равномерно). Пусть m = 200.

Для каждого момента времени t сгенерируйте соответствующее значение $y^{\text{ист}}$ при некоторых параметрах $a_2^{\text{ист}}, a_1^{\text{ист}}$ и $a_0^{\text{ист}}$. Отразите выбранные параметры в отчете.

Существует задача дать оценку параметрам a по результатам физических измерений значений $y^{\text{наб}}$. Пусть результаты измерений отличаются от истинных значений в силу действия случайной аддитивной помехи (случайность подчиняется нормальному закону распределения N(0, σ)). Сгенерируйте набор возможных $y^{\text{наб}}$. Отразите в отчете выбранное значение σ.

Для каждого набора параметров a определим m невязок:

$$\delta_i = y_i^{\text{наб}} - \mathrm{y}(t_i, a)$$

По сгенерированному набору точек $(t_i, y_i^{\text{наб}})$ дайте оценку параметрам закона с учетом знания общей формулы тремя различными способами:

- Пусть параметры $a_2^{\text{оц2}}, a_1^{\text{оц2}}$ и $a_0^{\text{оц2}}$ будут такими, что сумма квадратов невязок будет минимальна.
- Пусть параметры $a_2^{\text{оц1}}, a_1^{\text{оц1}}$ и $a_0^{\text{оц1}}$ будут такими, что сумма абсолютных значений невязок будет минимальна.
- Пусть параметры $a_2^{\text{оц}\infty}, a_1^{\text{оц}\infty}$ и $a_0^{\text{оц}\infty}$ будут такими, что максимальное абсолютное значение невязки будет минимально.

a) Постройте в одной координатной плоскости графики *y(t, a), y(t,* $a^{\text{оц2}}$*), y(t,* $a^{\text{оц1}}$*), y(t,* $a^{\text{оц}\infty}$*).*

b) Вычислите как отличается каждый из оценочных параметров от своего истинного значения. Как меняется это отличие при изменении σ?

c) Скорректируйте $y_0^{\text{наб}}$ и $y_{m-1}^{\text{наб}}$ пусть одно из них будет на 50 больше, а другое на 50 меньше. Постройте новые оценочные значения параметров $a^{\text{оц}}$ и соответствующие графики. Какая из оценок получилась более устойчивой к выбросам?