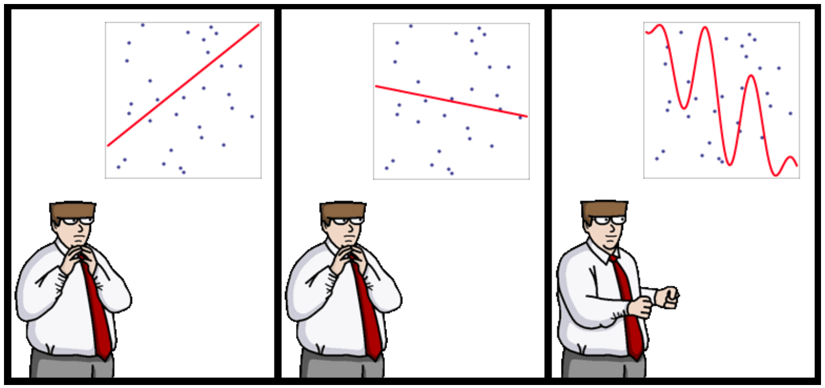
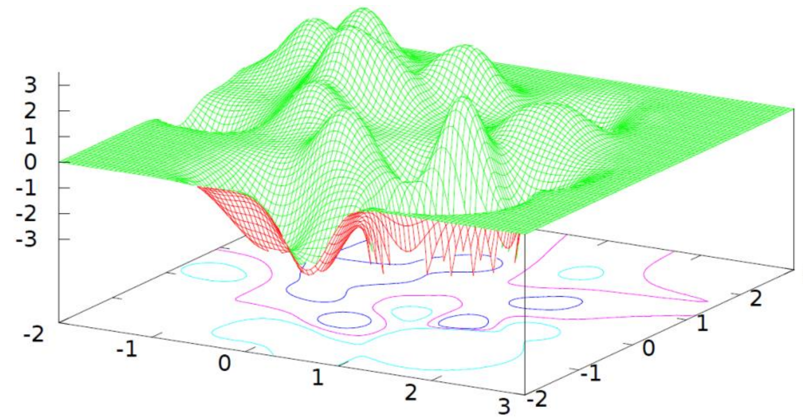


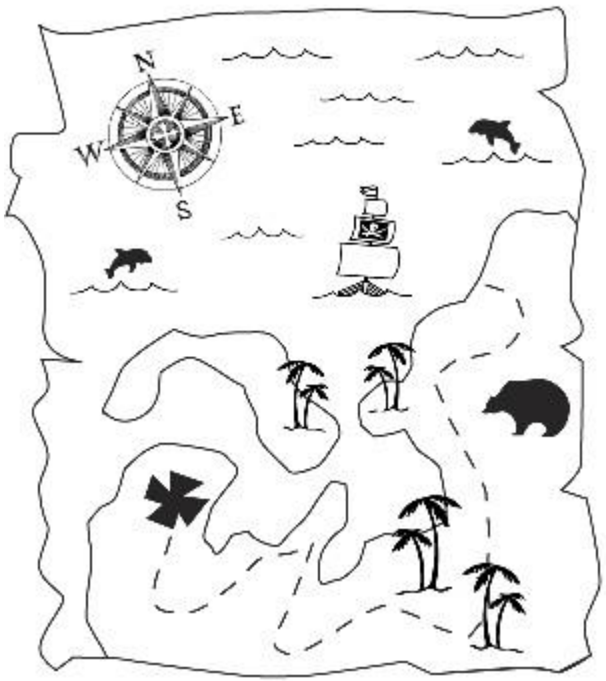


Анализ данных на практике

Оптимизационные задачи

Кантор Виктор

План



1. Как возникают задачи оптимизации
2. Градиентные методы оптимизации
3. Поиск глобального экстремума

1. Как возникают задачи оптимизации

Задача регрессии

$x_1, x_2, \dots, x_l$ - точки, в которых известны значения некоторой величины:

$$y_1, y_2, \dots, y_l$$

Мы строим прогнозирующий алгоритм:

$$y_i \approx a(x_i)$$

Но что значит «примерно равно»?

Задача регрессии

$x_1, x_2, \dots, x_l$ - точки, в которых известны значения некоторой величины:

$$y_1, y_2, \dots, y_l$$

Мы строим прогнозирующий алгоритм:

$$y_i \approx a(x_i)$$

Но что значит «примерно равно»?

Например, это:

$$\sum_{i=1}^{l} (y_i - a(x_i))^2 \rightarrow min$$

Задача регрессии

$x_1, x_2, \dots, x_l$ - точки, в которых известны значения некоторой величины:

$$y_1, y_2, \dots, y_l$$

Мы строим прогнозирующий алгоритм:

$$y_i \approx a(x_i)$$

Но что значит «примерно равно»?

В общем случае:

$$\sum_{i=1}^{l} L(y_i, a(x_i)) \quad \to min$$

Задача классификации

$x_1, x_2, \ldots, x_l$ - объекты, для которых известны их классы:

$$y_1, y_2, \ldots, y_l$$

Мы строим прогнозирующий алгоритм:

$$y_i = a(x_i)$$

Как выразить то, что он должен угадывать класс как можно чаще?

Задача классификации

$x_1, x_2, \dots, x_l$ - объекты, для которых известны их классы:

$$y_1, y_2, \dots, y_l$$

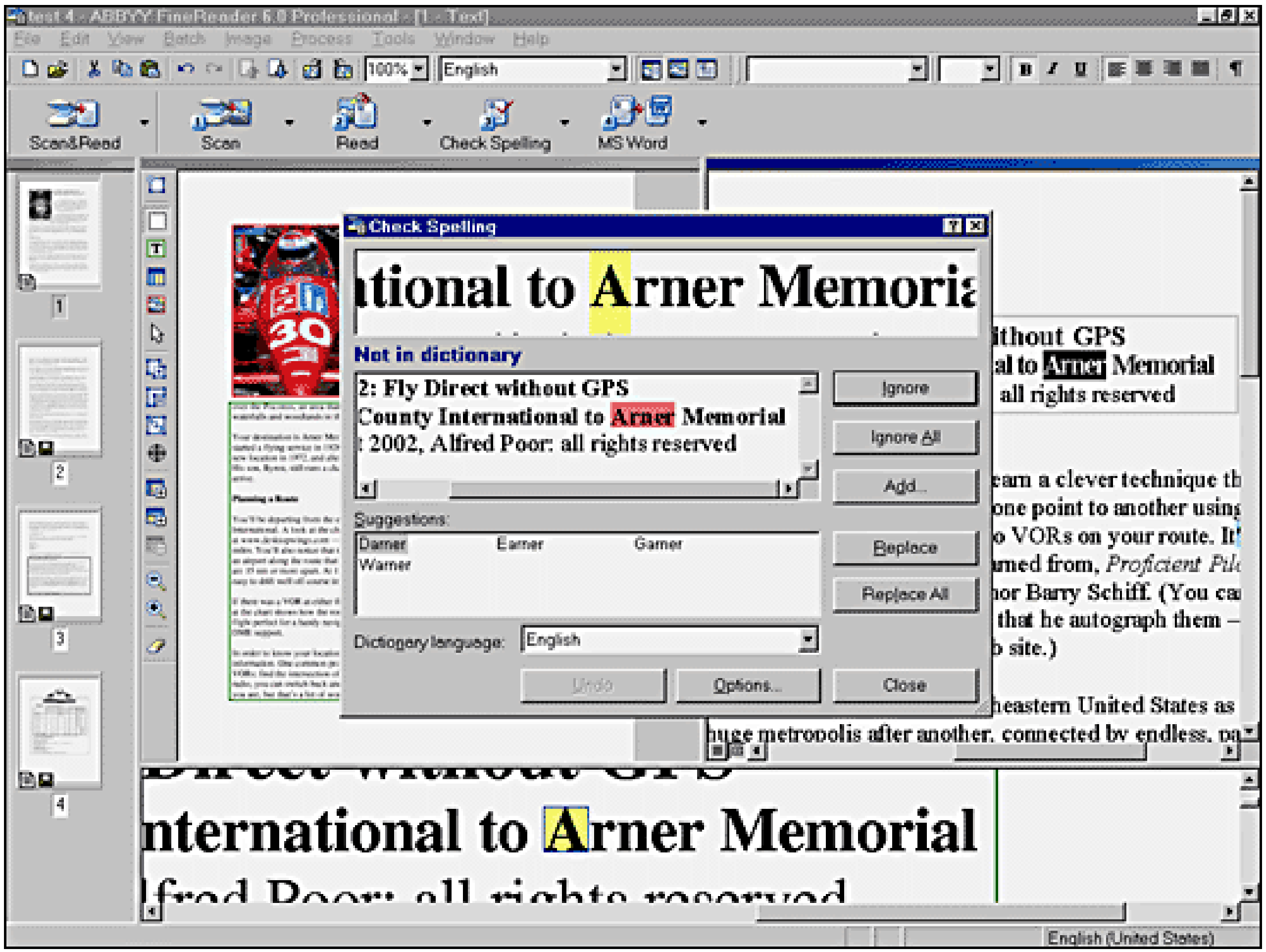
Мы строим прогнозирующий алгоритм:

$$y_i = a(x_i)$$

Как выразить то, что он должен угадывать класс как можно чаще?

$$\sum_{i=1}^{l} [y_i \neq a(x_i)] \to min$$

Сложный пример: исправление опечаток



Сложный пример: исправление опечаток

$$Suggest(w) = [w_1, w_2, \ldots, w_k]$$

В алгоритме есть параметры, которые когда-то были заданы «вручную». Хочется настроить их так, чтобы suggest был как можно «адекватней».

Есть выборка:
w (слово с опечаткой), cw(правильное написание)

Как сформулировать «адекватность» suggest’a, как настроить параметры?

Сложный пример: исправление опечаток

Возможное решение:

$$Suggest(w) = [w_1, w_2, \dots, w_k]$$
$$Pos\left(w_j, [w_1, w_2, \dots, w_k]\right) = j$$

$$\sum_{i=1}^{l} Pos(cw_i, Suggest(w_i)) \rightarrow min$$

Ещё пример: отбор признаков

Решается некоторая задача на наборе из N признаков

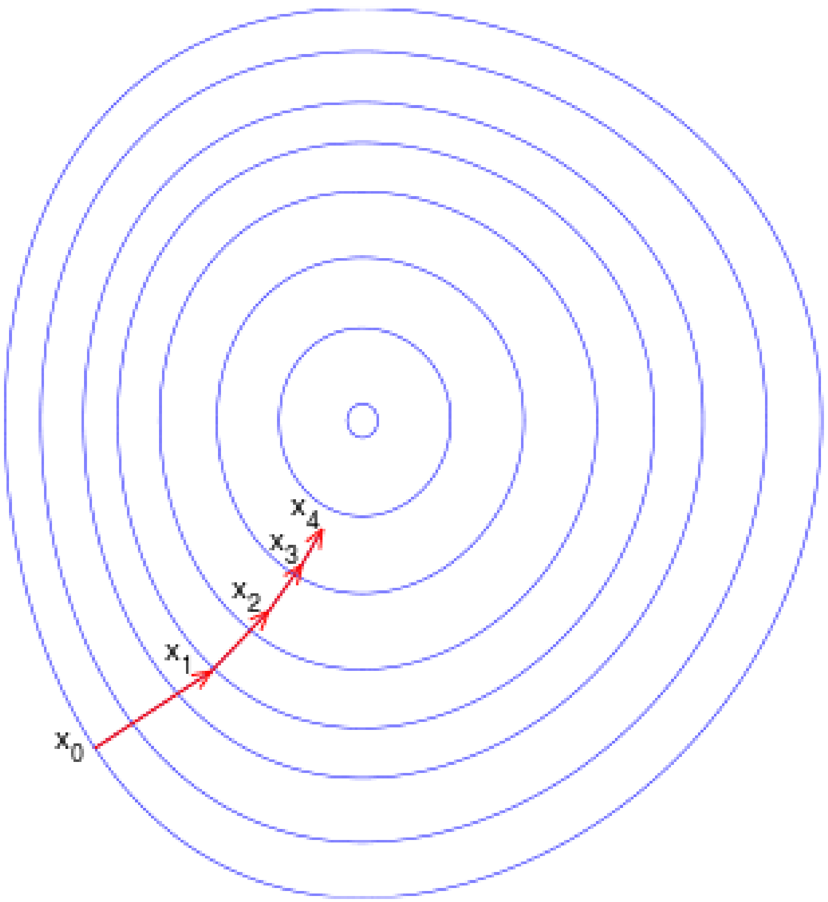
w – вектор длины N из 0 и 1

Q(w) – качество решения задачи, если алгоритм использует только те признаки, для которых в соответствующей координате w стоит 1

2. Градиентные методы оптимизации

Градиентный спуск

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n)$$



В задаче классификации/регрессии:

w – вектор параметров алгоритма

$$w_{n+1} = w_n - \gamma_n \sum_{i=1}^{l} \nabla L(x_i, y_i)$$

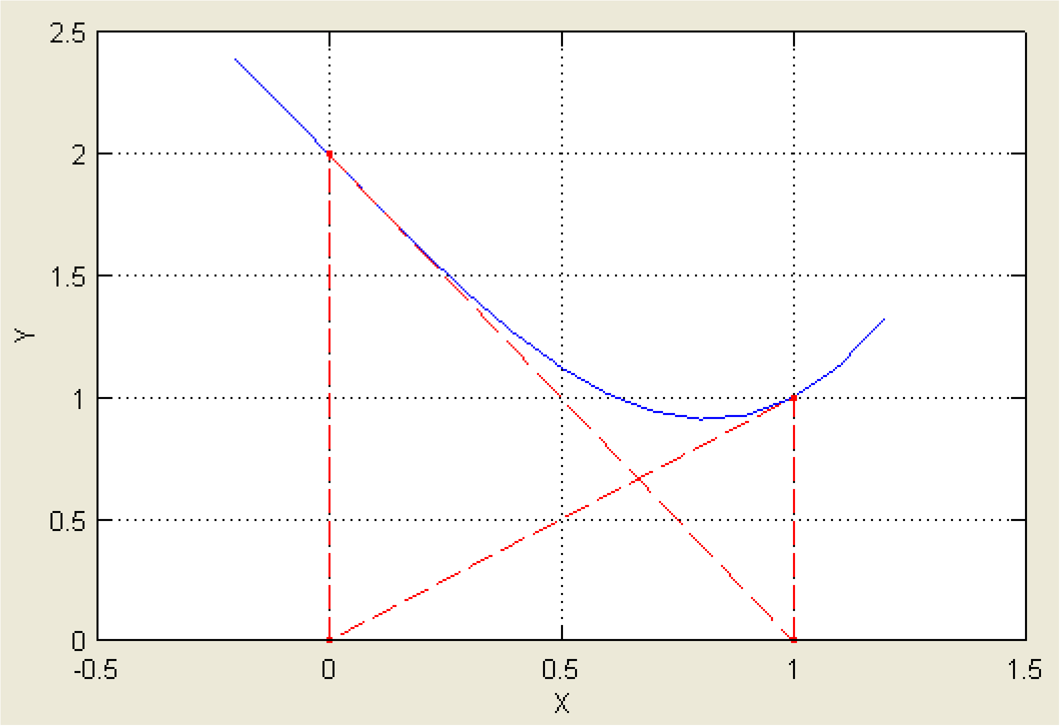
Стохастический градиентный спуск

$$w_{n+1} = w_n - \gamma_n \cancel{\sum_{i=1}^{l}} \nabla L(x_i, y_i)$$

$$\boxed{\begin{array}{c} w_{n+1} = w_n - \gamma_n \nabla L\,(x_i, y_i) \\ x_i - \text{случайный объект} \end{array}}$$

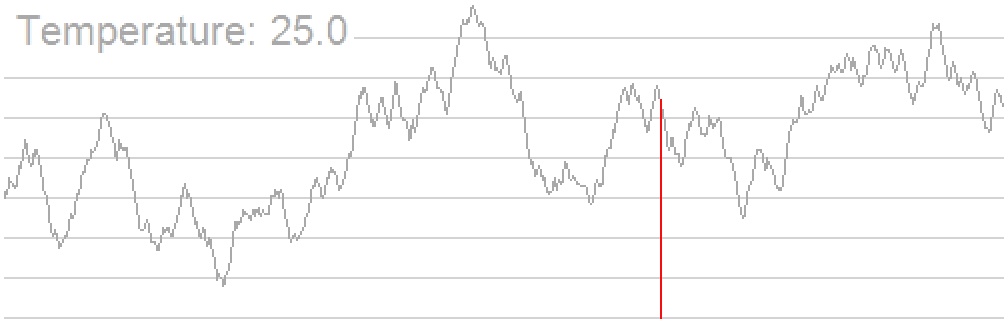
Метод Ньютона

Метод Ньютона: пример зацикливания



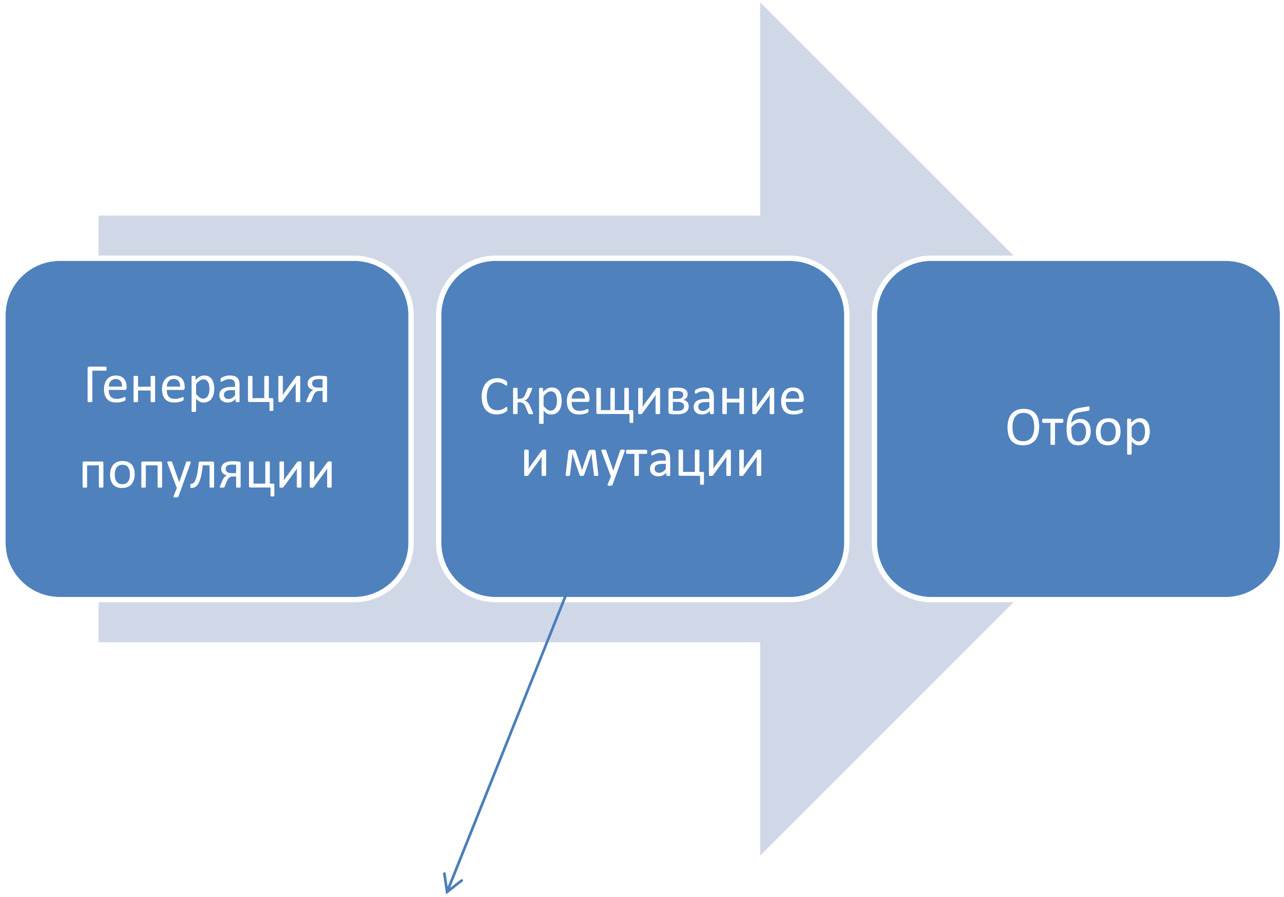
3. Поиск глобального максимума

Метод имитации отжига



$$P(\overline{x^*} \to \overline{x_{i+1}} \mid \overline{x_i}) = \begin{Bmatrix} 1, & F(\overline{x^*}) - F(\overline{x_i}) < 0 \\ \exp\left(-\dfrac{F(\overline{x^*}) - F(\overline{x_i})}{Q_i}\right), & F(\overline{x^*}) - F(\overline{x_i}) \geqslant 0 \end{Bmatrix}.$$

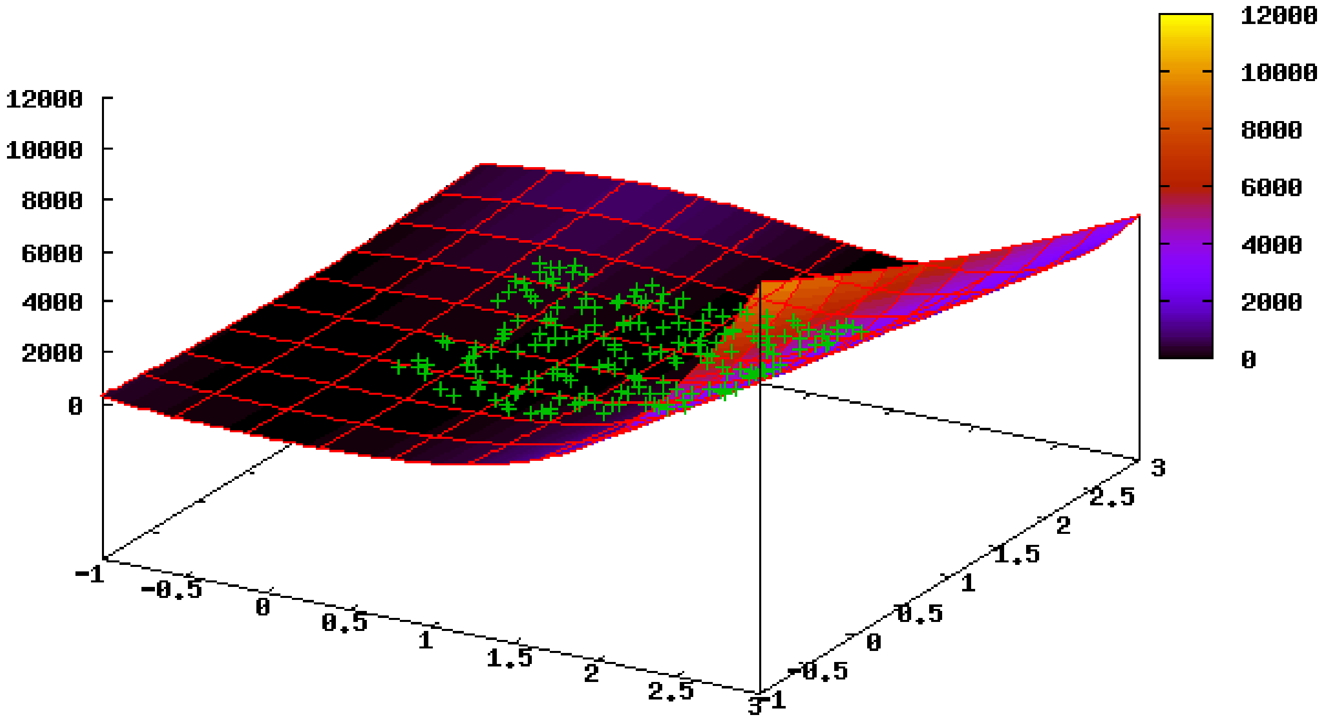
Дифференциальная эволюция



Каждую особь скрещиваем со случайной, но добавляя мутацию:

$$C'_j = C_j + F\,(A_j - B_j)$$

Демонстрация: дифф. эволюция

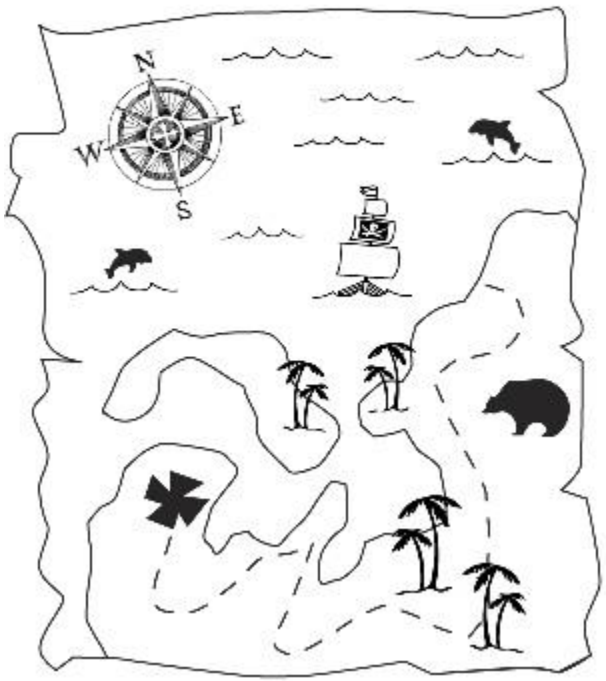


$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$ - функция Розенброка

Чем можно заняться уже сейчас

- Зарегистрироваться на kaggle.com и послать решения в учебные контесты (цифры и титаник)
- Ответить на вопросы:
  - Какие методы можно применить, когда нет готовой формулы для градиента оптимизируемой функции?
  - А когда функция негладкая или сильно зашумлена?
- Разобрать [введение в numpy, scipy, matplotlib]

План



1. Как возникают задачи оптимизации
2. Градиентные методы оптимизации
3. Поиск глобального экстремума

Спасибо за внимание!