

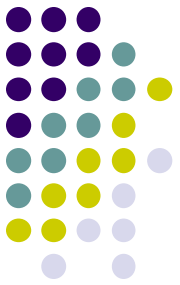
# Модификации ВР

---

Корлякова М.О.  
2019

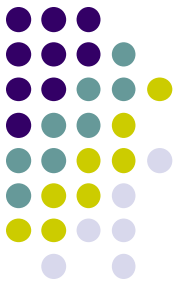


# Методы



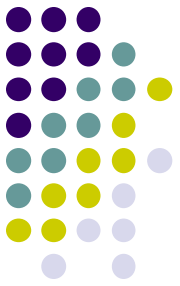
- Изменение способа определения шага
- Изменение способа вычисления  $\Delta\omega$

# Изменение способа определения шага

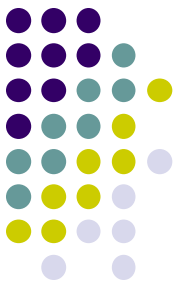


- Определять скорость обучения отдельно для каждого элемента  $W$ .
- Изменять скорость обучения для каждой итерации.
  - Если производная функции стоимости имеет постоянный знак для нескольких итераций, то скорость обучения растёт
  - Если производная функции стоимости показывает изменение знака для нескольких итераций, то скорость обучения растёт
- Вносить импульсные вариации в скорость обучения.

# Изменение способа вычисления $\Delta\omega$



- Градиентный спуск
- Методы сопряженных градиентов.
- Квази-Ньютоновские методы.
- Метод стохастического градиента



# Градиентный спуск

- Вектор  $\nabla f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$  частных производных функции  $f(X) = f(x_1, x_2, \dots, x_n)$
- Задаёт направление наискорейшего возрастания этой функции
- $$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial E(t)}{\partial w_{ij}(t)}$$

# Метод сопряженных градиентов



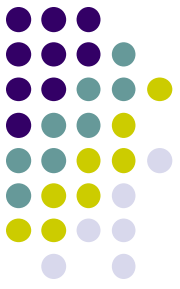
- Задача минимизации квадратичной функции.
- Функцию стоимости аппроксимируем квадратичной функцией
$$f(w) = \frac{1}{2} W^T H W - \partial E_{av}(W) / \partial W \quad W + c$$
- В качестве направления наискорейшего спуска выбираем резидуальную ошибку  $r(n)$  – ошибка нахождения минимума
- Не использует Гессиан, но его приближение

# Нелинейный алгоритм сопряжен градиентов



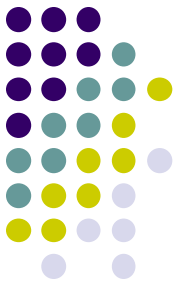
- Инициализация
- Вычисления
  1.  $W(0)$  по ВР находим градиент  $g(0)$
  2.  $S(0)=r(0)=-g(0)$  ( $r(0)$  – резидуальная ошибка)
  3. Для шага  $n$  - линейный поиск параметра  $\eta(n)$ 
    - Находится группа (которая является нетривиальным интервалом), гарантированно содержащая минимум.
    - Разделение на подгруппы меньшего размера

# Нелинейный алгоритм сопряжен градиентов



4. Ошибка  $r(n) < e\_max(n)$ , если да, то закончить обучение.  $e\_max(n) = r(0)/10000$
5.  $W(n+1) = W(n) + \eta(n) s(n)$
6. По ВР находим  $g(n+1)$
7.  $r(n+1) = -g(n+1)$
8. Вычисляем  $b$  по Полаку-Рибьеру  
 $b(n+1) = \max\{0, (r^T(n+1)(r(n+1) - r(n)))/(r^T(n) r(n))\}$
9. Изменяем значение направления
10.  $s(n+1) = r(n+1) + b(n+1) s(n)$
11.  $n = n+1$  и к 3.



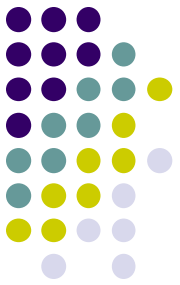


# Метод Ньютона

- Формула ньютона

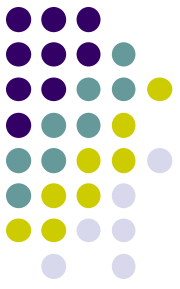
$$\mathbf{W}^{k+1} = \mathbf{W}^k - \mathbf{H}^{-1}(\mathbf{W}^k) \cdot \text{grad} E(\mathbf{W}^k),$$

- $\mathbf{H}(\mathbf{W}^k)$  Гессиан функции ошибки



# LM

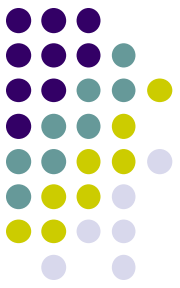
- $\Delta W = -(J^T J + I * M)^{-1} J^T E$
- $H \approx J^T J$
- J- Якобиан
- $g = J^T E$



# BFGS

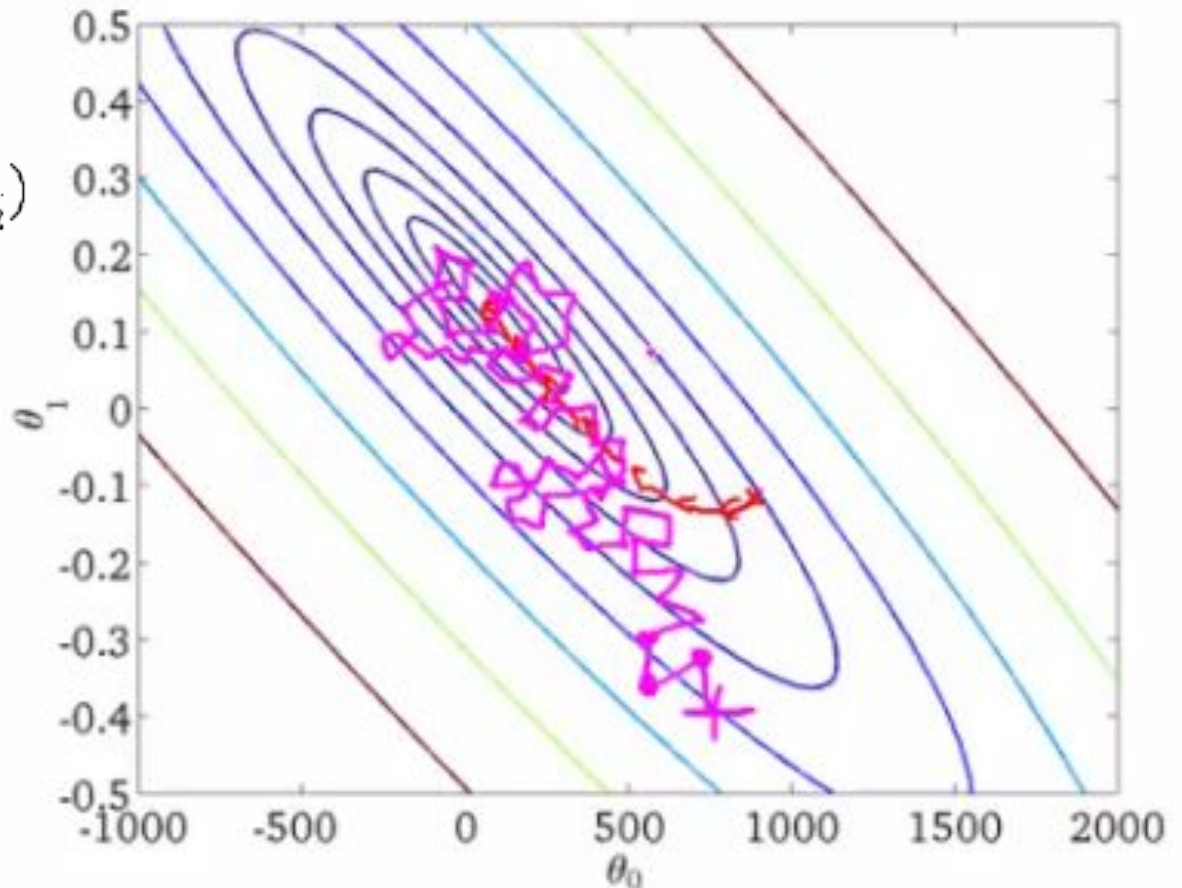
- $\Delta W = -H^{-1}g(n)$
  - $s(n)P(n) = -g(n)$
  - $s(n) = s(n-1) + u(n)$
  - $V(n) = W(n+1) - W(n)$
  - $Y(n) = g(n+1) - g(n)$
  - $S(n+1)V(n) = Y(n)$
  - Broyden-Fletcher-Goldfarb-Shanno
- $$P(n+1) = -g(n) + [(V(n)^* g(n+1))^* s(n) / (Y(n)^* s(n))]$$

# Стохастический градиентный спуск

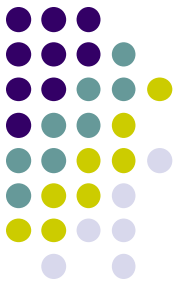


- SGD (Stochastic Gradient Descent)

$$Q := \sum_{i=1}^I L(a(x_i, w), y_i)$$



# Стохастический градиентный спуск



## 1. Текущий шаг

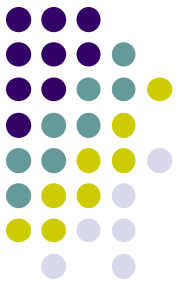
- Выбрать объект  $x_i$  из (например, случайным образом);
- Вычислить выходное значение алгоритма и ошибку;  $\epsilon_i := L(a(x_i, w), y_i)$
- Сделать шаг градиентного спуска;

$$w := w - \eta L'_a(a(x_i, w), y_i) \varphi'(\langle w, x_i \rangle) x_i$$

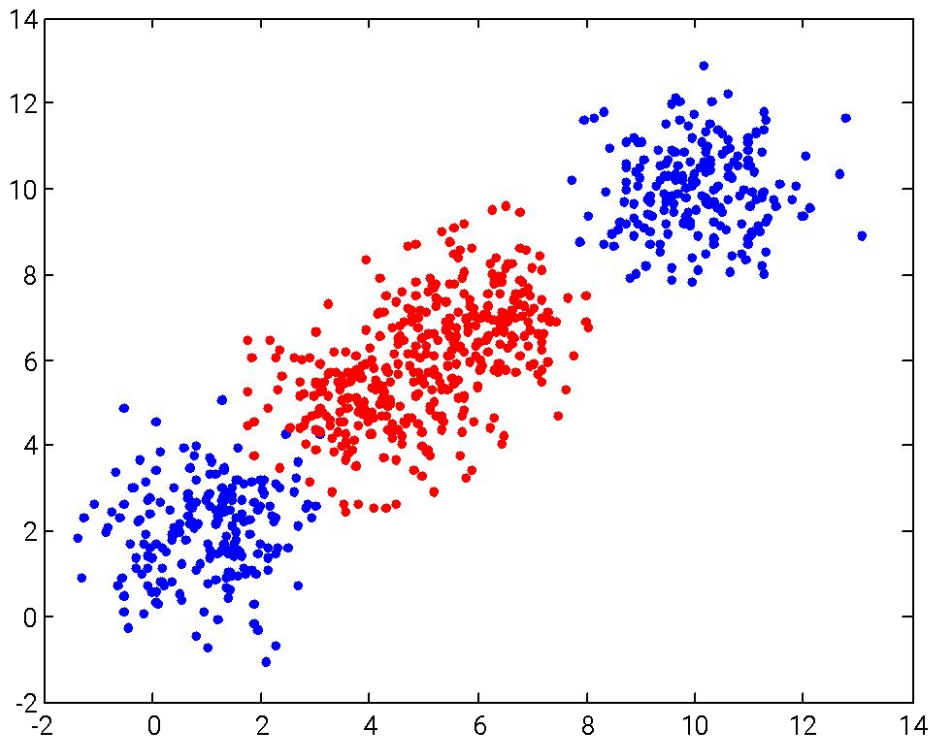
- Оценить значение функционала;  $Q := (1 - \lambda)Q + \lambda \epsilon_i$

2. Пока значение не стабилизируется и/или веса не перестанут изменяться.

# Изменение весов



- 2 класса



# Изменение весов

