Проксимальный градиентный метод для обучения моделей с L1-регуляризацией

Практикум на ЭВМ 2017/2018

Фоминская Галина Евгеньевна

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

14 ноября 2017 г.

Градиентный спуск

- $F(\omega) \to \min_{\omega}$
- $F(\omega)$ – гладкая и выпуклая

Градиентный спуск

- $F(\omega) \to \min\limits_{\omega}$
- $F(\omega)$ – гладкая и выпуклая
- Решение:

$$\omega^{(k+1)} = \underline{\omega^{(k)} - \eta \nabla F(\omega^{(k)})}$$

Градиентный спуск

- $F(\omega) \to \min\limits_{\omega}$
- $F(\omega)$ – гладкая и выпуклая
- Решение:

$$\omega^{(k+1)} = \underline{\omega^{(k)} - \eta \nabla F(\omega^{(k)})}$$

- Заметим, что это эквивалентно:

$$\omega^{(k+1)} = \arg\min_{\omega} \frac{1}{2} \|\omega - (\underline{\omega^{(k)} - \eta \nabla F(\omega^{(k)})})\|_2^2$$

Постановка задачи

- $F(\omega) = f(\omega) + g(\omega) \to \min\limits_{\omega}$
- $f(\omega)$ гладкая и выпуклая
- $g(\omega)$ выпуклая, негладкая, простая

Проксимальный метод

$$\omega^{(k+1)} = \arg\min_{\omega} \left(\frac{1}{2\eta} \|\omega - (\omega^{(k)} - \eta \nabla f(\omega^{(k)}))\|_2^2 + g(\omega) \right)$$

Проксимальный метод

$$\omega^{(k+1)} = \arg\min_{\omega}\left(\frac{1}{2\eta}\|\omega - (\omega^{(k)} - \eta\nabla f(\omega^{(k)}))\|_2^2 + g(\omega)\right)$$

$$= \arg\min_{\omega}\left(f(\omega^{(k)}) + \nabla f(\omega^{(k)})^T(\omega - \omega^{(k)}) + \frac{1}{2\eta}\|\omega - \omega^{(k)}\|_2^2 + g(\omega)\right)$$

Проксимальный метод

- $\omega^{(k+1)} = \arg\min\limits_{\omega} \left(\frac{1}{2\eta} \|\omega - \underline{(\omega^{(k)} - \eta \nabla f(\omega^{(k)}))}\|_2^2 + g(\omega) \right)$

Проксимальный метод

- $\omega^{(k+1)} = \arg\min_{\omega}\left(\frac{1}{2\eta}\|\omega - \underline{(\omega^{(k)} - \eta\nabla f(\omega^{(k)}))}\|_2^2 + g(\omega)\right)$
- $prox(z) = \arg\min_{\omega}\left(\frac{1}{2\eta}\|\omega - z\|_2^2 + g(\omega)\right)$

Проксимальный метод

- $\omega^{(k+1)} = \arg\min_{\omega} \left(\frac{1}{2\eta} \|\omega - \underline{(\omega^{(k)} - \eta \nabla f(\omega^{(k)}))}\|_2^2 + g(\omega) \right)$
- $prox({\color{blue}z}) = \arg\min_{\omega} \left(\frac{1}{2\eta} \|\omega - {\color{blue}z}\|_2^2 + g(\omega) \right)$
- $\omega^{(k+1)} = prox(\underline{\omega^{(k)} - \eta \nabla f(\omega^{(k)})})$

$L1-$регуляризация

В нашем случае:

$$F(\omega) = f(\omega) + \underbrace{\alpha\|\omega\|_1}_{g(\omega)} \to \min_{\omega}$$

$$\omega^{(k+1)} = prox(\omega^{(k)} - \eta \nabla f(\omega^{(k)}))$$

$$prox(z) = \arg\min_{\omega} \left(\frac{1}{2\eta} \|\omega - z\|_2^2 + \alpha\|\omega\|_1 \right)$$

Решение

- $prox(z) = \sum_{i=1}^{d} prox(z_i)$
- Можно вычислять покомпонентно:
  $prox(z_i) = \arg\min_{\omega_i} \left(\frac{1}{2\eta} (\omega_i - z_i)^2 + \alpha |\omega_i| \right)$

Решение

- $prox(z) = \sum_{i=1}^{d} prox(z_i)$
- Можно вычислять покомпонентно:
  $prox(z_i) = \arg\min_{\omega_i} \left(\frac{1}{2\eta}(\omega_i - z_i)^2 + \alpha|\omega_i| \right)$
- $prox(z_i) = \begin{cases} z_i - \alpha\eta, & z_i > \alpha\eta \\ 0, & |z_i| < \alpha\eta \\ z_i + \alpha\eta, & z_i < -\alpha\eta \end{cases}$

Soft Thresholding