

Некоторые замечания

гуу

Оптимизация

Оптимизация

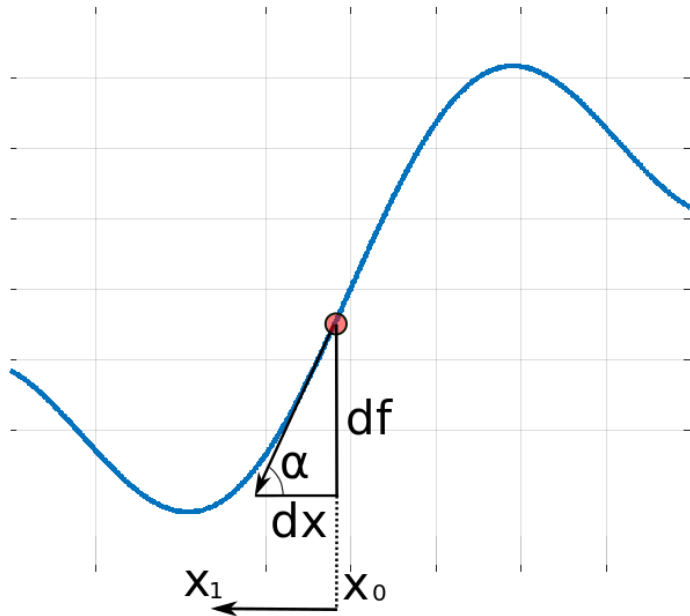
Перекрестная энтропия

Линал

Градиентный спуск

```
>>edit showGrad  
>>peaks(0.1)  
>>peak(0.05)
```

$$\Delta x = -\eta \left. \frac{df}{dx} \right|_{x=x_0} \quad x_1 = x_0 + \Delta x_0$$



1. Подбор скорости обучения
2. Остановка в локальном минимуме
3. Затухание (взрыв) градиента

Градиентный спуск: скорость обучения

1. Momentum
2. RMSProp
3. Adam

Momentum

$$\theta_{i+1} = \theta_i - \eta \left. \frac{df}{d\theta} \right|_{\theta=\theta_i} + \gamma (\theta_i - \theta_{i-1})$$

RMSProp (root mean square propagation)

$$E[g^2]_i = \gamma E[g^2]_{i-1} + (1 - \gamma) g_i^2, \text{ где } g \equiv \frac{df}{d\theta}$$

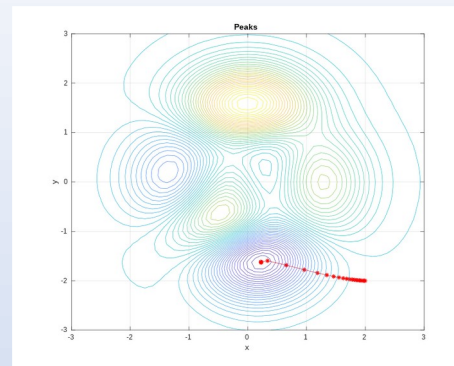
$$\theta_{i+1} = \theta_i - \frac{\eta}{\sqrt{E[g^2]_i + \epsilon}} g_i$$

Adam (adaptive moment estimation)

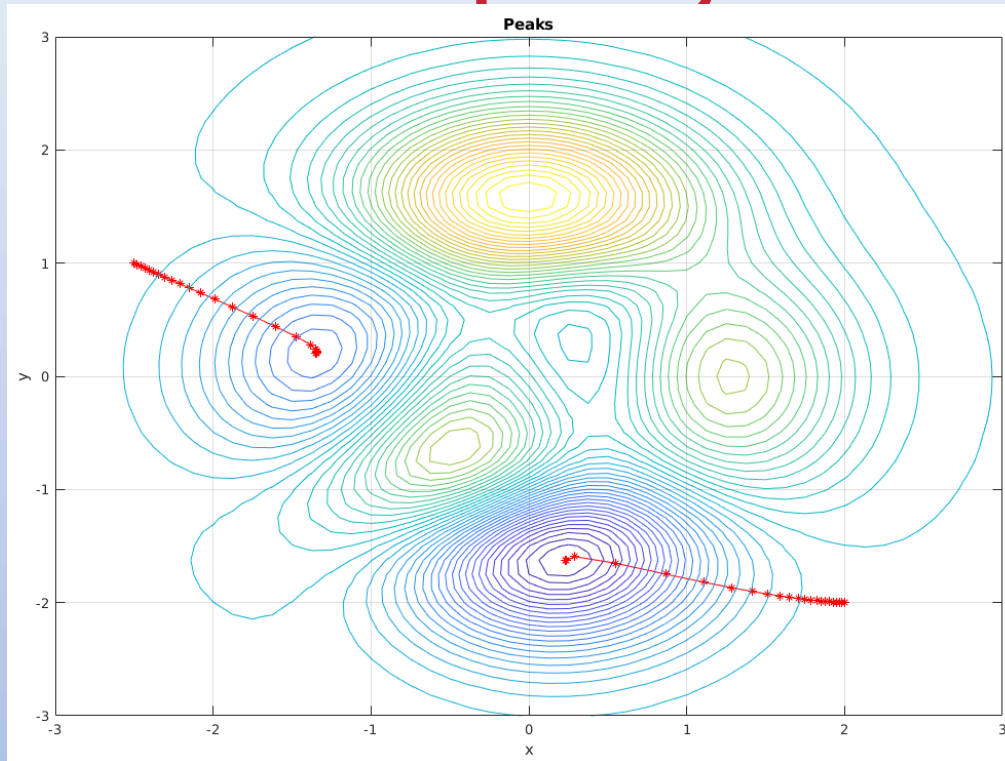
$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) g_i$$

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) g_i^2$$

$$\theta_{i+1} = \theta_i - \frac{\eta}{\sqrt{v_i + \epsilon}} m_i$$



Градиентный спуск: поиск глобального экстремума



Информационная энтропия

$$i = \log_2 N$$

$$i = \sum \frac{1}{N} \log_2 N = - \sum \frac{1}{N} \log_2 \frac{1}{N}$$

N — мощность алфавита

$$H(x) = - \sum_{i=1}^n p_i(x) \log_2 p(x)$$

Перекрестная энтропия

Расстояние Кульбака - Лейблера

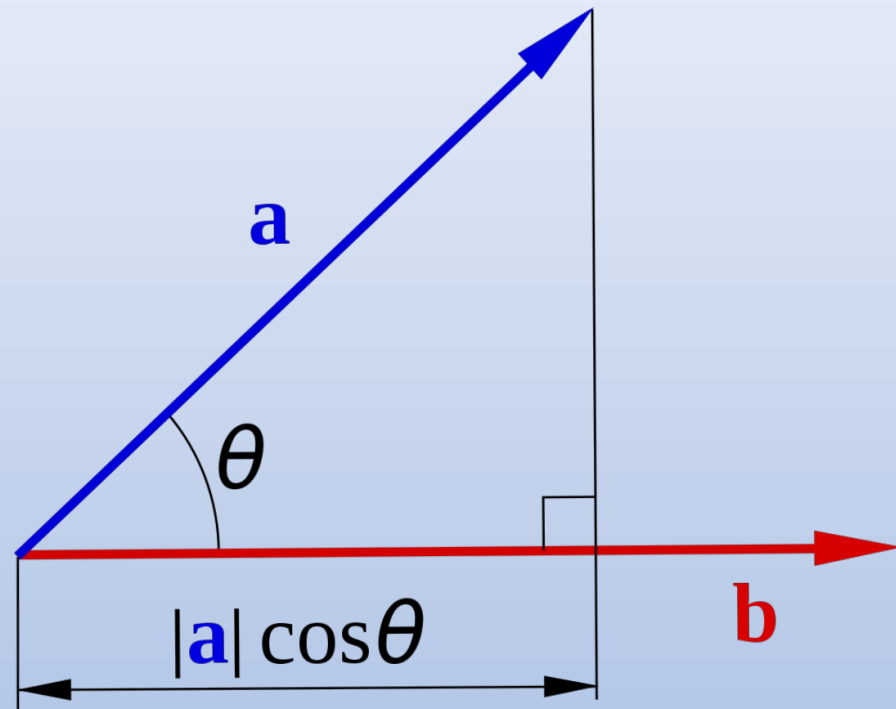
$$D_{KL}(P \parallel Q) = \int_X p \log \frac{p}{q} d\mu$$

$$D_{KL}(P \parallel Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$$

$$D_{KL}(P \parallel Q) = \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n p_i \log q_i = H(p) - H(p, q)$$

Линейная алгебра

$$(a, b) = |a| |b| \cos(\theta)$$



Линейная алгебра

$$(a, b) = |a| |b| \cos(\theta)$$

