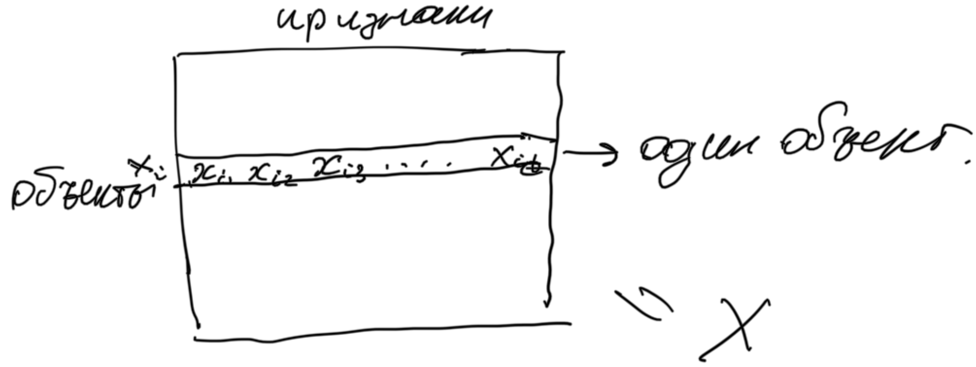
$$\frac{1}{l}\sum_{i=1}^{l}\left(\langle w, x_i\rangle - y_i\right)^2 \to \min_w$$

$$\|$$

$$\frac{1}{l}\|Xw - y\|^2 \to \min_w$$



$= X$

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

$$x_{i1} w_1 + x_{i2} w_2 + \dots$$

$$x_{id} w_d = \langle x_i, w\rangle$$

$$Xw = \begin{pmatrix} \langle w, x_1\rangle \\ \langle w, x_2\rangle \\ \vdots \\ \langle w, x_l\rangle \end{pmatrix}$$

$$Xw - y = \begin{pmatrix} \langle w, x_1\rangle - y_1 \\ \dots \end{pmatrix}$$

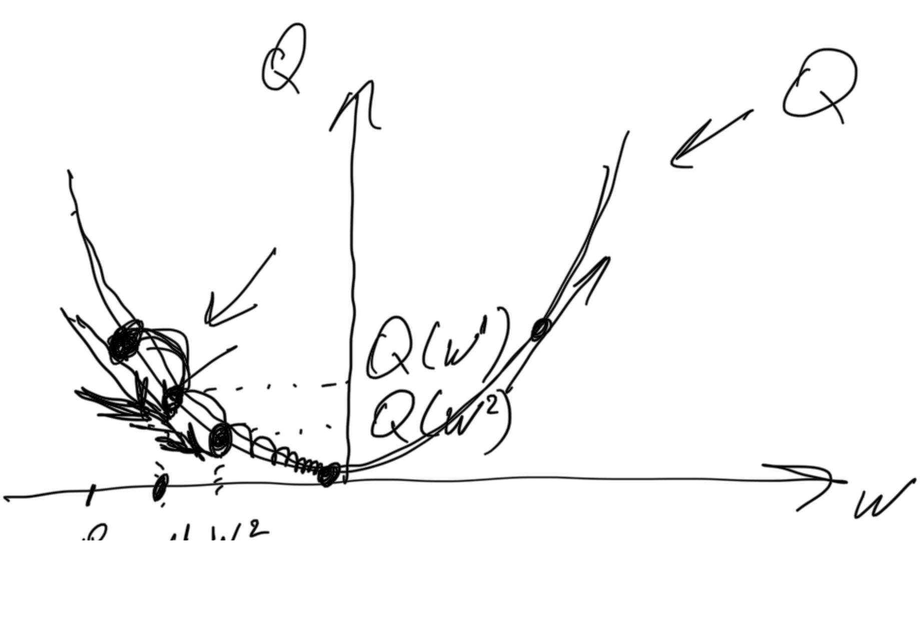
$$\begin{pmatrix} \langle w, x_2 \rangle - y_2 \\ \vdots \\ \langle w, x_l \rangle - y_l \end{pmatrix}$$

$$\underline{\frac{1}{l}\sum_{i=1}^{l}\left(\langle w, x_i \rangle - y_i\right)^2} \to \min_{w}$$

$$\underline{w} = \underline{(X^T X)}^{-1} X^T \underline{y}.$$

$$\nabla f(x_1, \dots, x_n) = \left(\frac{\partial f}{\partial x_i}\right)_{i=1}^{n}$$

$-\nabla f.$



$$\boxed{Q(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}\left(\langle w, x_i\rangle - y_i\right)^2 = \frac{1}{\ell}\sum_{i=1}^{\ell}\mathcal{L}(y_i, a(x_i))}$$

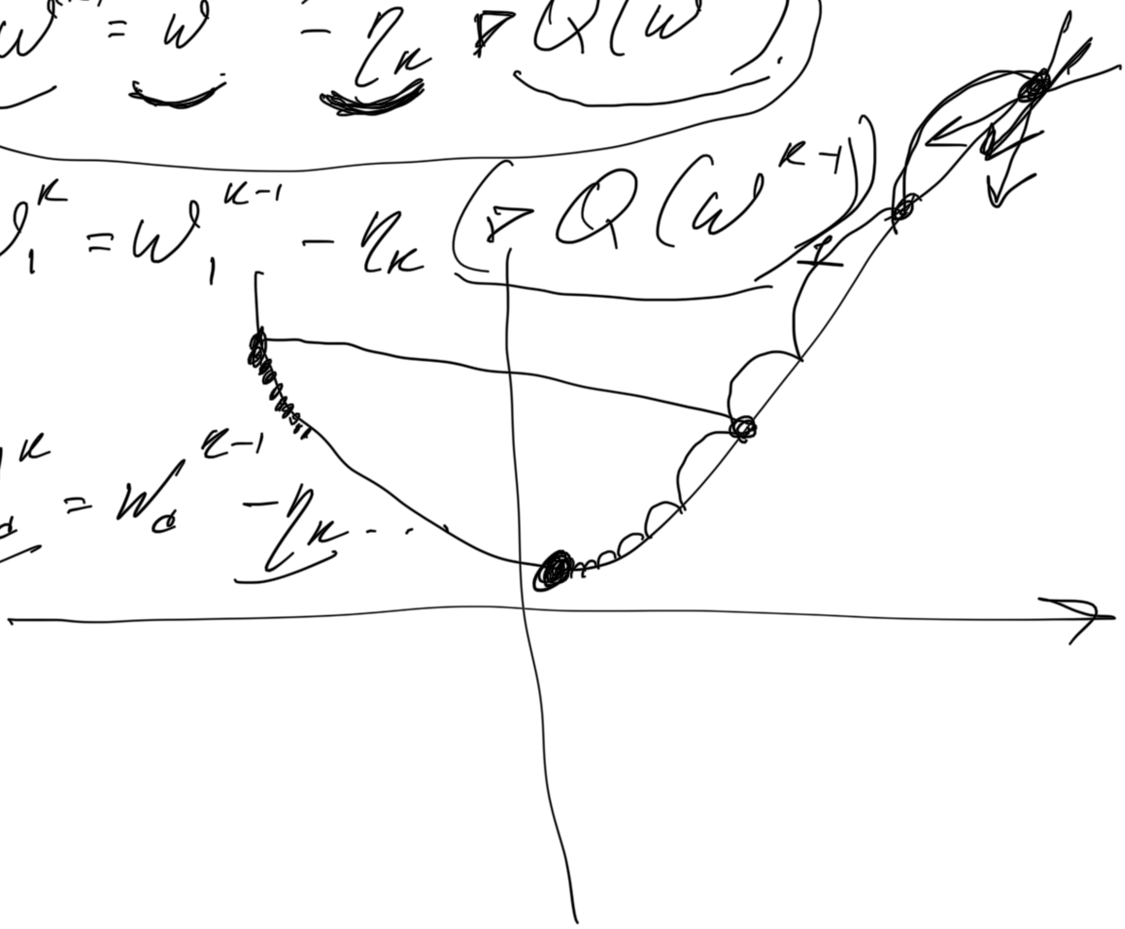
$\boxed{GD}$ $w^0 = \text{random}$

$$w^{(k)} = w^{(k-1)} - \eta_k \nabla Q(w^{k-1})$$

$$w_1^k = w_1^{k-1} - \eta_k \nabla Q(w^{k-1})$$

$\vdots$

$$w_d^k = w_d^{k-1} - \eta_k \ldots$$



$$\eta_k = \frac{1}{k}$$

1) $\|\nabla Q(w^k)\| \le \varepsilon \quad 10^{-5}$

2) $\|w^k - w^{k-1}\| < \varepsilon$

2) $\| w - w \| < \varepsilon$

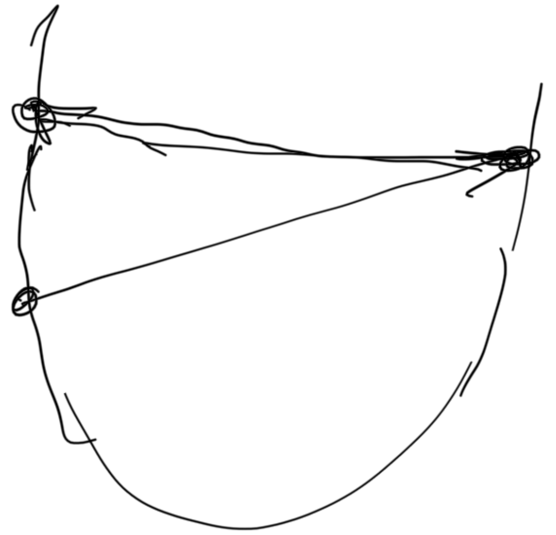
3) $\| Q(w^k) - Q(w^{k-1}) \| < \varepsilon$

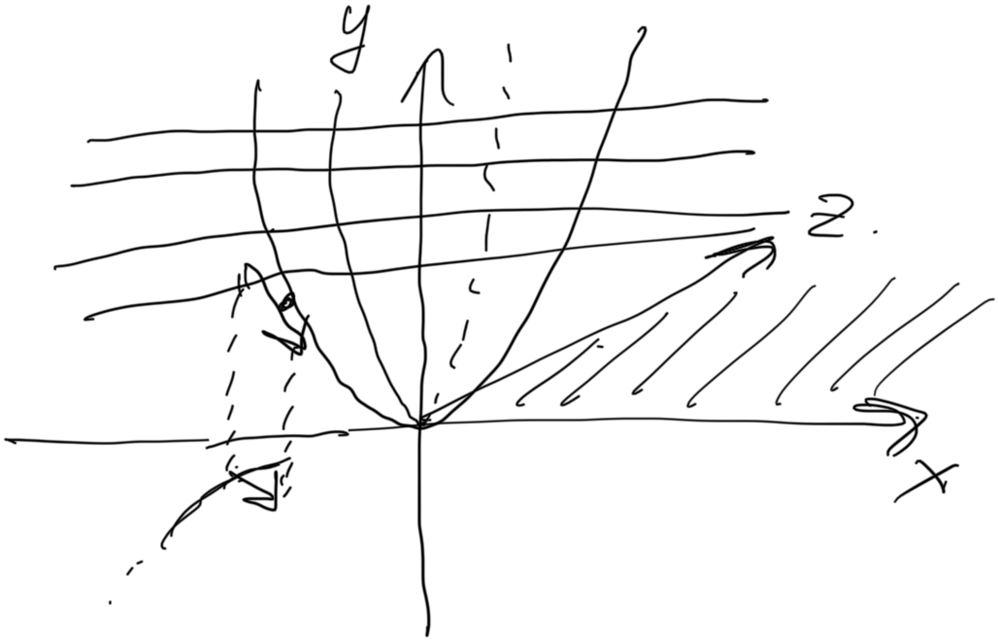
$$\nabla Q(w) = \frac{1}{\ell} \sum_{i=1}^{\ell} \nabla \mathcal{L}(y_i, a(x_i))$$

<u>SGD</u>

$$\nabla_w Q(w) \approx \nabla_w \mathcal{L}(y_i, a(x_i))$$

$$w^k = w^{k-1} - \eta_k \nabla \mathcal{L}(y_i, a(x_i))$$



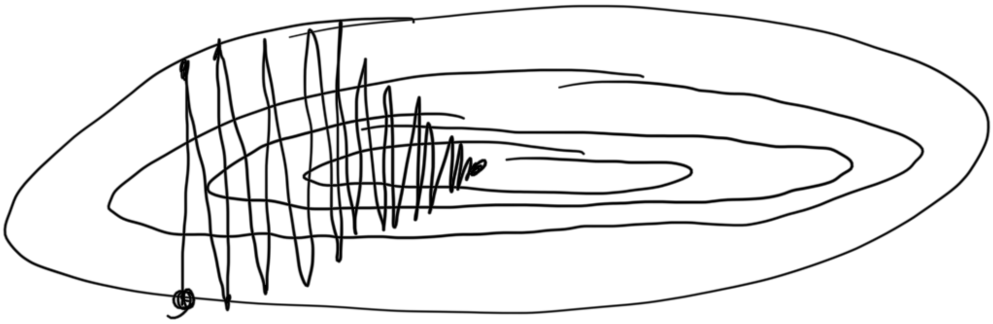


$$\nabla Q(w) \approx \frac{1}{m} \sum_{i=1}^{m} \nabla \mathcal{L}(y_i, a(x_i))$$

$$Q(w_1, w_2)$$

$N+1$

Метод инерции (momentum)



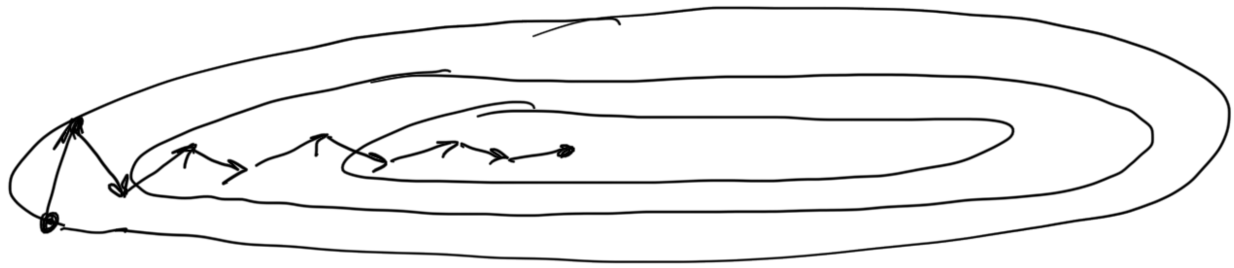
$$h_0 = 0$$

$$h_k = \alpha h_{k-1} + \eta_k \nabla Q(w^{k-1})$$

$$h_1 = \eta_k \nabla Q(w^0)$$

$$h_2 = \alpha h_1 + \eta_k \nabla Q(w^1)$$

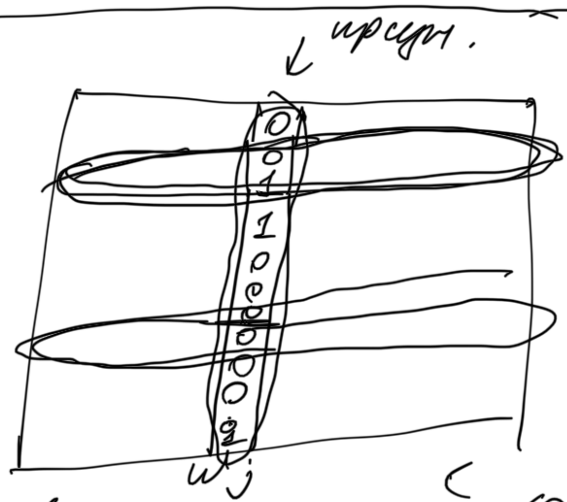
$$\boxed{w^k = w^{k-1} - h_k}$$



AdaGrad

$(w_0, w_1, \ldots, w_d)$

$Q(w)$



$$\nabla Q(w) = \frac{1}{\ell} \sum_{i=1}^{\ell} \nabla \mathcal{L} = \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{\partial \mathcal{L}}{\partial w_1}, \ldots, \frac{\partial \mathcal{L}}{\partial w_d} \right)$$

SGD

$\approx w_j^0$

$w_j, w_j, \ldots \ldots w_j$

$k = 100$

$\boxed{\eta_k = \frac{1}{k}}$ $\quad \frac{1}{100}$

$= 0 \quad 7^0$

$G \; = G \quad + \left(\nabla_{w} Q(w^{k-1}) \right)^2$

$$w_j^{(k)} = w_j^{k-1} - \frac{\eta_k}{\sqrt{G_{kj} + \varepsilon}} \left(\nabla_w Q(w^{k-1})\right)_j$$

RMSprop

$$G_{kj} = \alpha G_{k-1,j} + (1-\alpha)\left(\nabla_w Q(w^{k-1})\right)_j^2$$

Adam = Adagrad + momentum