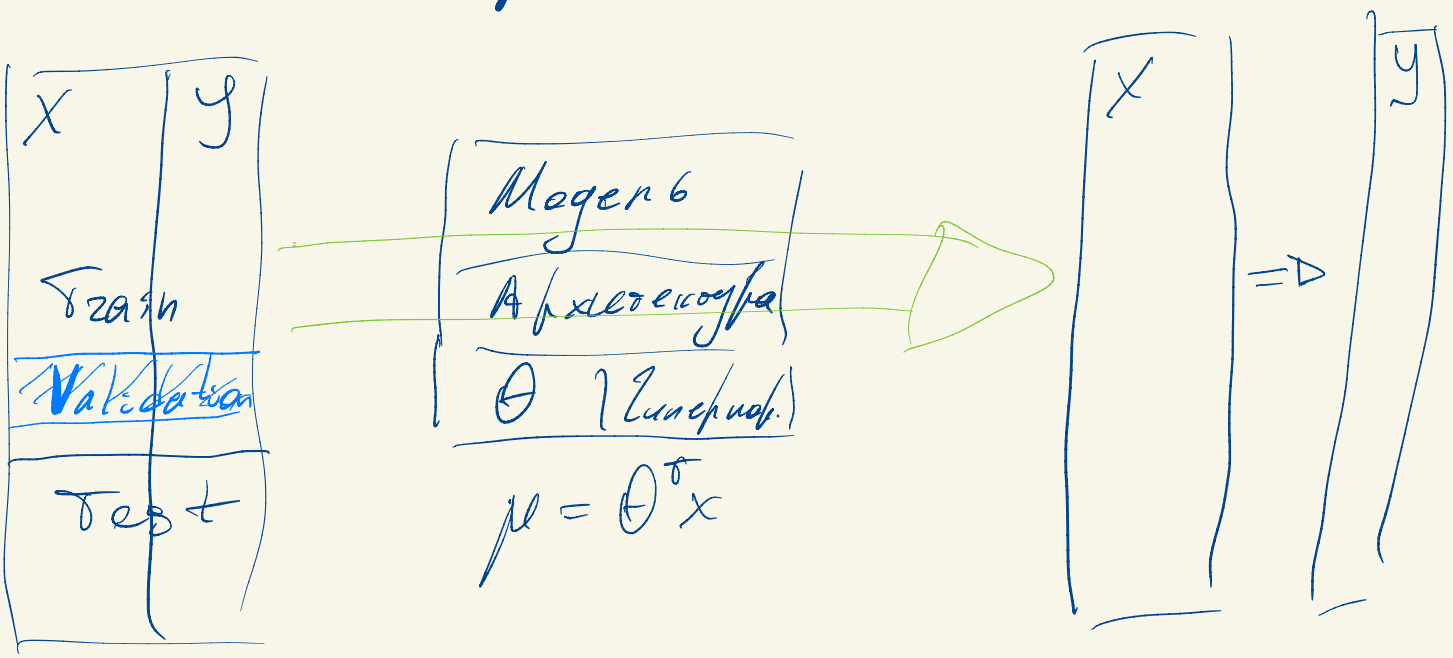
Supervised learning



$M = 0,7 \pm \Delta M$

Bootstrap

Градиентная оптимизация

$\mu = \theta^T x$ $\quad x \Leftrightarrow y$

$\theta = ?$

$$\mathcal{L} = (y - \theta^T x)^T (y - \theta^T x)$$

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \theta x_i)^2$$

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta, X, y)$$ — „обучение"

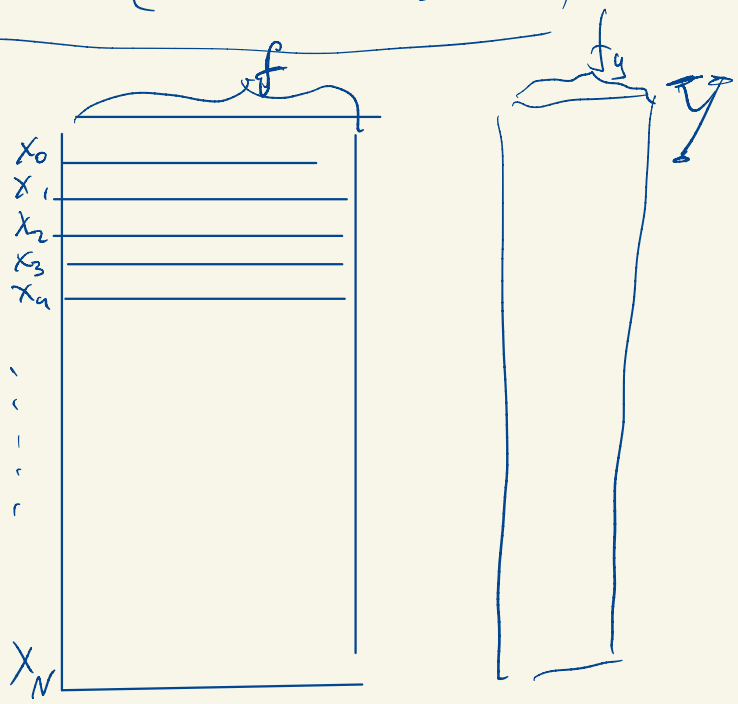
Как?

$\nabla_\theta \mathcal{L}(\theta^*) = 0$ $\quad \frac{\partial^2 \mathcal{L}}{\partial \theta^2} - ?$

$$\nabla_\theta \mathcal{L} = -2 X^T y + 2 X^T X \theta = 0$$

$$\theta^* = (X^T X)^{-1} X^T y$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2} = 2 X^T X$$



$$\theta^* = (X^T X)^{-1} X^T y$$

$X : N \times f$

$y : N \times f_y$

$X^T y : f \times N \cdot N \times f_y$

$f \times f_y$

$X^T X \quad f \times N \cdot N \times f$

$f \times f$

$f \times f \cdot f \times f_y$

$\theta^* : f \times f_y$

$X\theta \quad N \times f \cdot f \times f_y$

$N \times f_y$

$$\theta^* = (X^T X)^{-1} X^T y$$

$z = \theta^T x \qquad p = \frac{1}{1+e^{-z}} = \sigma(z)$

$= \sigma(\theta^T x)$

$$\theta^* = \underset{\theta}{\arg\min}\, \mathcal{L}(\theta, X, Y)$$

Методы оптимизации

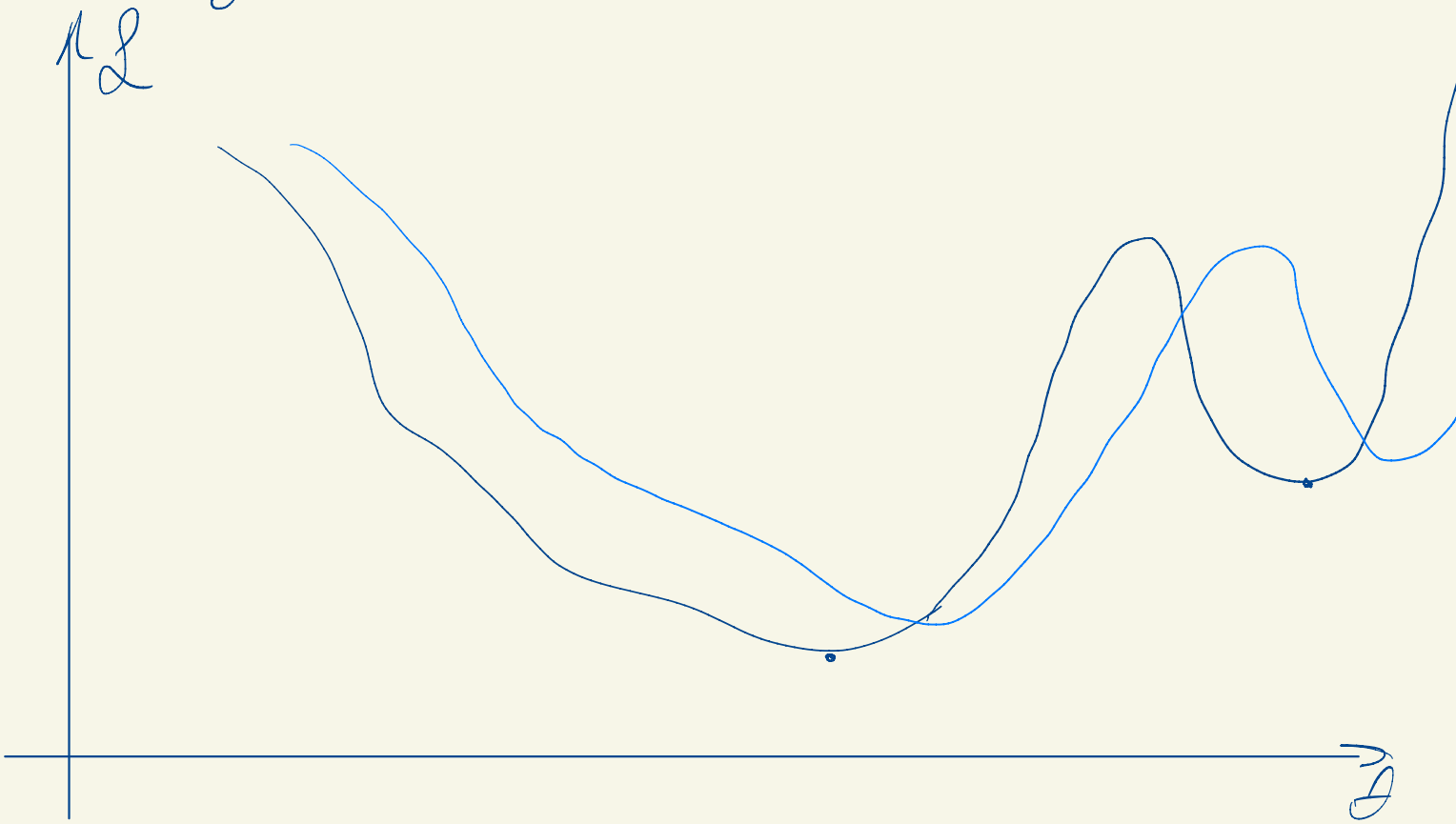
1. Метод Ньютона
2. Эволюционные
3. Градиентные

$\mathcal{L}(\theta, \bar{X}, \bar{Y})$ $\quad X\theta^T = \bar{Y}$ $\quad |f \times f_y|$

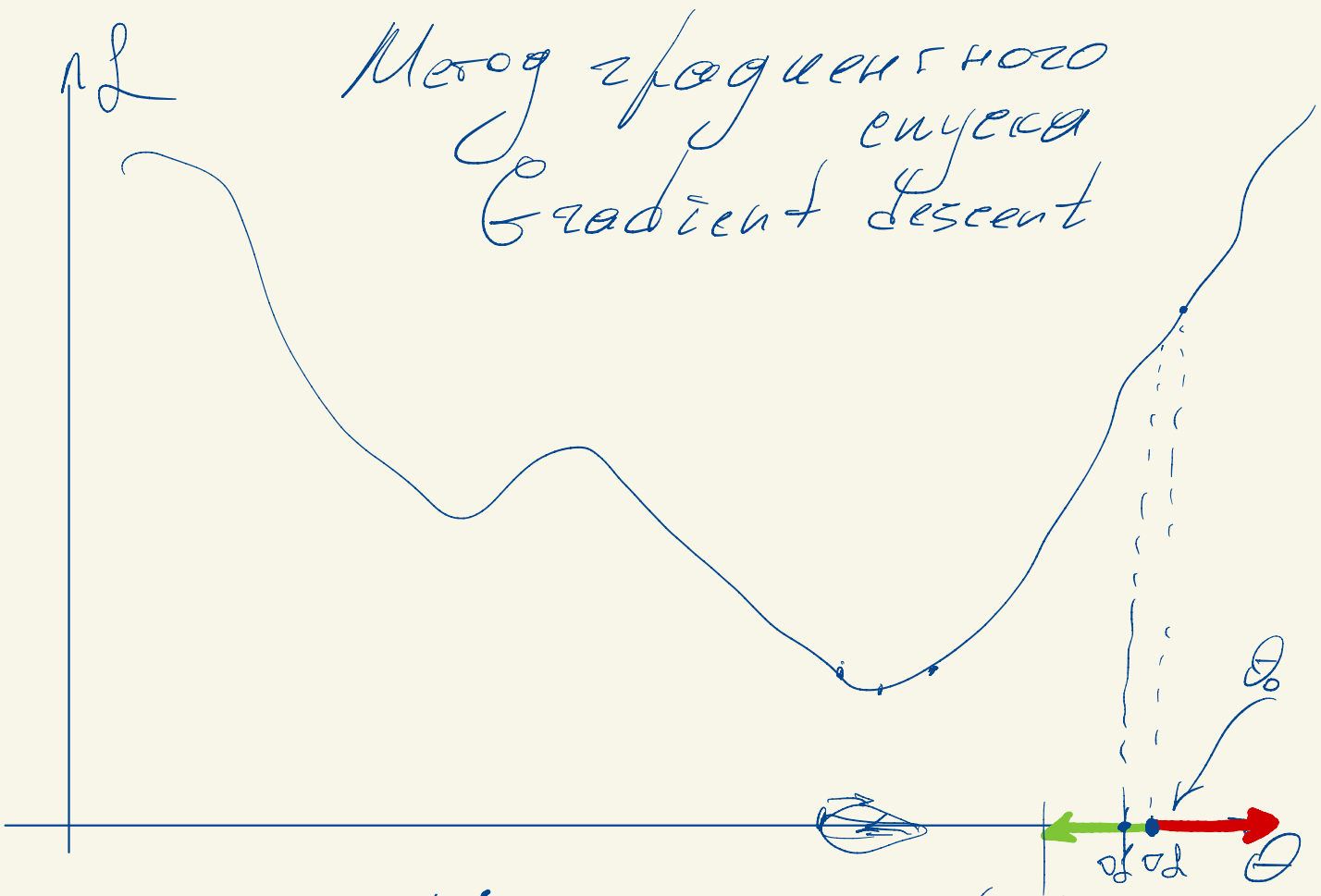
$$\nabla_\theta \mathcal{L} = \vec{a}_{\theta_0} \frac{\partial \mathcal{L}}{\partial \theta_0} + \vec{a}_{\theta_1} \frac{\partial \mathcal{L}}{\partial \theta_1} + \dots + \vec{a}_{\theta_d} \frac{\partial \mathcal{L}}{\partial \theta_d}$$

$\nabla_\theta \mathcal{L}$ - указывает направление наискорейшего роста f-ии $\mathcal{L}$

$-\nabla_\theta \mathcal{L}$ — " — — " — — " —
убывания f-ии $\mathcal{L}$



Метод градиентного спуска
Gradient Descent



0) θ_0 ; $t = 0$; $\varepsilon = 10^{-6}$

1) $\nabla_\theta L$

2) $\theta_{t+1} = \theta_t - \eta \nabla_\theta L$

3) ?$(|\nabla_\theta L| \le \varepsilon)$ Stop

θ_0 - начальное приближение

$\nabla_\theta L$ - градиент L по θ

η - шаг обучения (маленькое число $\eta << 1$)

t - номер итерации оптимизации

$\Rightarrow \theta^*$