Методы восстановления плотности распределения

Евгений Борисов

Восстановление плотности

$$a(x) = \underset{y \in Y}{argmax}\, \lambda_y P(y)\, p(x|y)$$

Байесовский классификатор

λ_y - потеря для объектов y

P(y) - априорная вероятность класса y
(доля примеров класса y,
пропорция классов должна соответствовать)

p(x|y) - ф-ция правдоподобия класса y (плотность)

Восстановление плотности

подходы к оценке плотности распределения:

параметрический подход

$$\hat{p}(x)=\varphi(x,\theta)$$

НЕпараметрический подход

$$\hat{p}(x)=\frac{1}{mV(h)}\sum_{j=1}^{m}K\left(\frac{\rho(x,x_j)}{h}\right)$$

смеси распределений

$$\hat{p}(x)=\sum_{j=1}^{k}w_j\varphi_j(x,\theta_j)$$

Восстановление плотности

смеси распределений

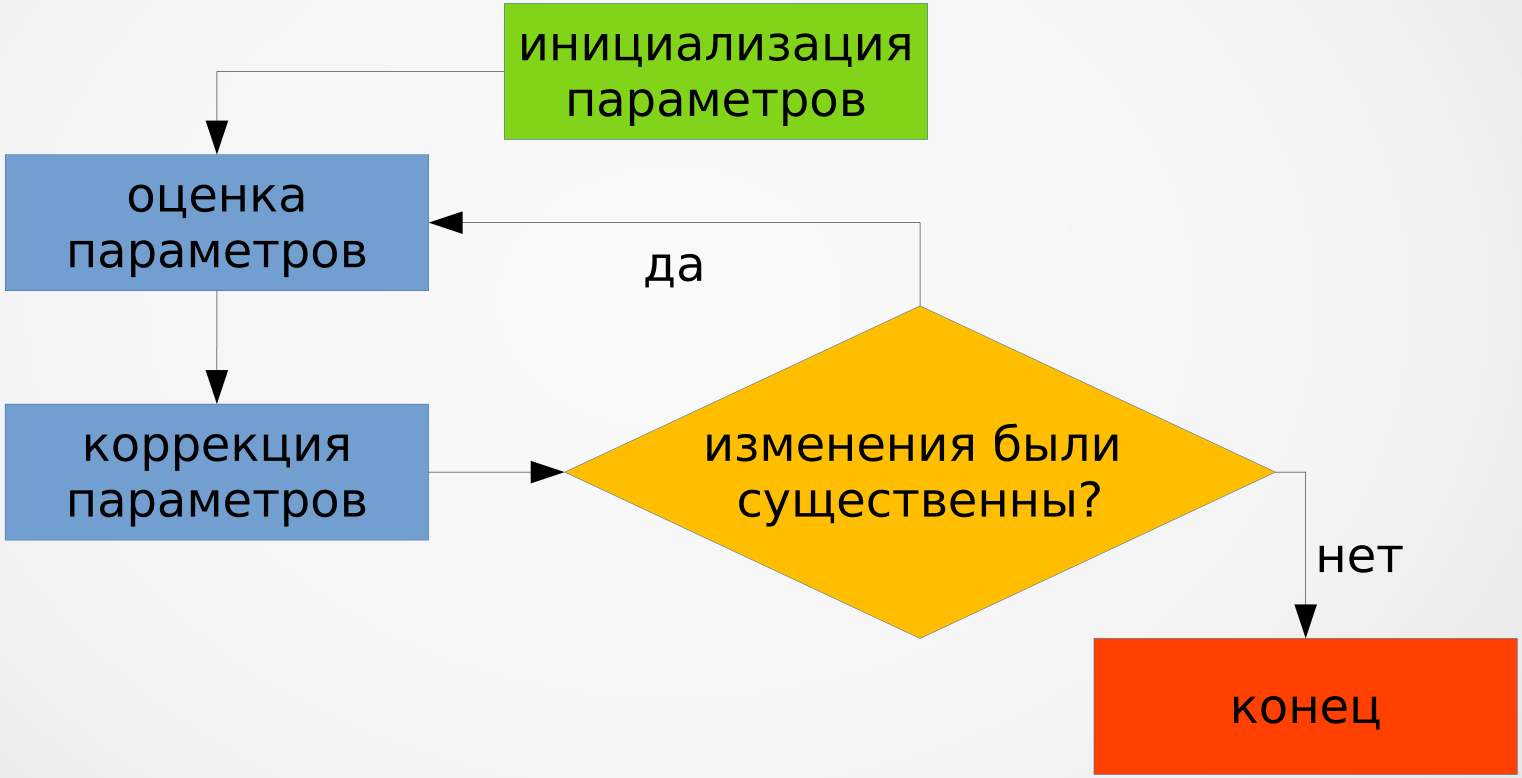
$$\hat{p}(x)=\sum_{j=1}^{k} w_j \varphi_j(x, \theta_j);$$

$$\sum_{j=1}^{k} w_j = 1; \qquad w_j \geq 0$$

Восстановление плотности

EM (expectation–maximization algorithm):

базовый вариант алгоритма



Восстановление плотности

EM (expectation–maximization algorithm)

оценка

$$g_{ij} = \frac{w_j \varphi_j(x_i, \theta_j)}{\sum_{k=1}^{s} w_k \varphi_k(x_i, \theta_k)}$$

i=1...m
m - количество примеров X
s - количество компонент смеси

Восстановление плотности

EM (expectation–maximization algorithm)

<u>оценка</u>

$$g_{ij}=\frac{w_j\,\varphi_j(x_i,\theta_j)}{\sum_{k=1}^{s} w_k\,\varphi_k(x_i,\theta_k)}$$

<u>коррекция</u>

$$w_j=\frac{1}{m}\sum_{i=1}^{m} g_{ij}$$

$$\theta_j=\underset{\theta}{argmax}\sum_{i=1}^{m} g_{ij}\ln\varphi_j(x_i,\theta)$$

i=1...m
m - количество примеров X
s - количество компонент смеси

Восстановление плотности

EM (expectation–maximization algorithm)

<u>оценка</u>

$$g_{ij}=\frac{w_j\,\varphi_j\left(x_i,\theta_j\right)}{\sum_{k=1}^{s} w_k\,\varphi_k\left(x_i,\theta_k\right)}$$

<u>коррекция</u>

$$w_j=\frac{1}{m}\sum_{i=1}^{m} g_{ij}$$

$$\theta_j=\underset{\theta}{argmax}\sum_{i=1}^{m} g_{ij}\ln\varphi_j\left(x_i,\theta\right)$$

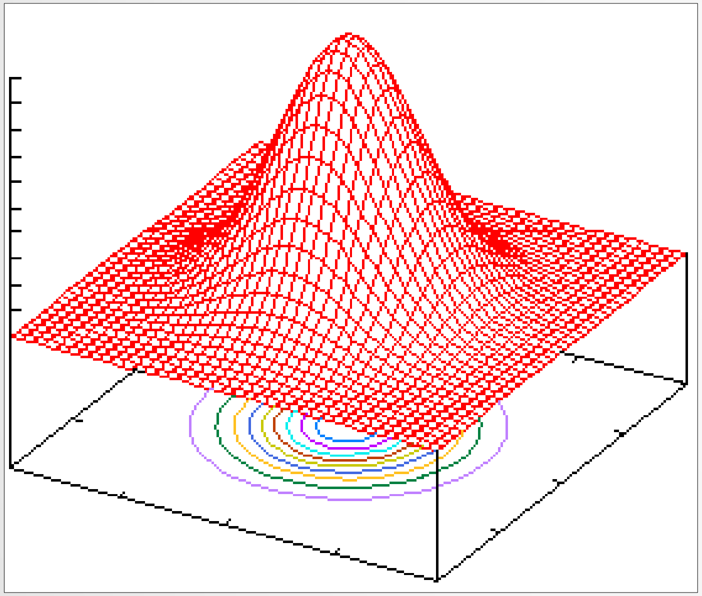
i=1...m
m - количество примеров X
s - количество компонент смеси

<u>условие остановки:</u> параметры не изменились

$$\left|g_{ij}(t-1)-g_{ij}(t)\right|<\delta\,;0<\delta<1$$

Восстановление плотности

параметрический подход:
допустим - **p(x)** это нормальная *n-мерная* плотность



$$p(x)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n\det\Sigma}}$$

Восстановление плотности

EM для нормальной плотности

$$p(x)=\sum_k w_k N(x;\Sigma_k,\mu_k)$$

n-мерная гауссовская плотность

$$p(x)=N(x;\Sigma,\mu)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n det\,\Sigma}}$$

оценки параметров для максимального правдоподобия имеют следующий вид

мат.ожидание

$$\hat{\mu}=\frac{1}{m}\sum_{i=1}^{m}x_i$$

матрица ковариаций

$$\hat{\Sigma}=\frac{1}{m}\sum_{i=1}^{m}(x_i-\hat{\mu})(x_i-\hat{\mu})^T$$

Восстановление плотности

EM для нормальной плотности $$p(x)=\sum_k w_k N(x;\Sigma_k,\mu_k)$$

$$N(x;\Sigma,\mu)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n \det\Sigma}}$$

оценка: $$g_{ij}=\frac{w_j N(x_i;\Sigma_j,\mu_j)}{\sum_k w_k N(x_i;\Sigma_k,\mu_k)}$$

задача: $$\Sigma_j,\mu_j=\underset{\Sigma,\mu}{argmax}\sum_{i=1}^{m} g_{ij}\ln N(x_i;\Sigma,\mu)$$

Восстановление плотности

EM для нормальной плотности $p(x)=\sum_k w_k N(x;\Sigma_k,\mu_k)$

$$N(x;\Sigma,\mu)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n \det\Sigma}}$$

оценка: $$g_{ij}=\frac{w_j N(x_i;\Sigma_j,\mu_j)}{\sum_k w_k N(x_i;\Sigma_k,\mu_k)}$$

задача:

$$\Sigma_j,\mu_j=\underset{\Sigma,\mu}{argmax}\sum_{i=1}^{m} g_{ij}\ln N(x_i;\Sigma,\mu)$$

коррекция:

$$w_j=\frac{1}{m}\sum_{i=1}^{m} N_j \qquad N_j=\sum_{i=1}^{m} g_{ij}$$

вес компоненты

Восстановление плотности

EM для нормальной плотности $p(x)=\sum_k w_k N(x;\Sigma_k,\mu_k)$

оценка: $g_{ij}=\dfrac{w_j N(x_i;\Sigma_j,\mu_j)}{\sum_k w_k N(x_i;\Sigma_k,\mu_k)}$

$$N(x;\Sigma,\mu)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n \det\Sigma}}$$

задача: $\Sigma_j,\mu_j=\underset{\Sigma,\mu}{argmax}\sum_{i=1}^{m} g_{ij}\ln N(x_i;\Sigma,\mu)$

коррекция:

$$w_j=\frac{1}{m}\sum_{i=1}^{m} N_j$$

вес компоненты

$$N_j=\sum_{i=1}^{m} g_{ij}$$

$$\mu_j=\frac{1}{N_j}\sum_{i=1}^{m} g_{ij}x_i$$

мат.ожидание компоненты

Восстановление плотности

EM для нормальной плотности

$$p(x)=\sum_k w_k N(x;\Sigma_k,\mu_k)$$

оценка:

$$g_{ij}=\frac{w_j N(x_i;\Sigma_j,\mu_j)}{\sum_k w_k N(x_i;\Sigma_k,\mu_k)}$$

$$N(x;\Sigma,\mu)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^n \det\Sigma}}$$

задача:

$$\Sigma_j,\mu_j=\underset{\Sigma,\mu}{argmax}\sum_{i=1}^{m} g_{ij}\ln N(x_i;\Sigma,\mu)$$

коррекция:

$$w_j=\frac{1}{m}\sum_{i=1}^{m} N_j$$

$$N_j=\sum_{i=1}^{m} g_{ij}$$

$$\mu_j=\frac{1}{N_j}\sum_{i=1}^{m} g_{ij}x_i$$

вес компоненты

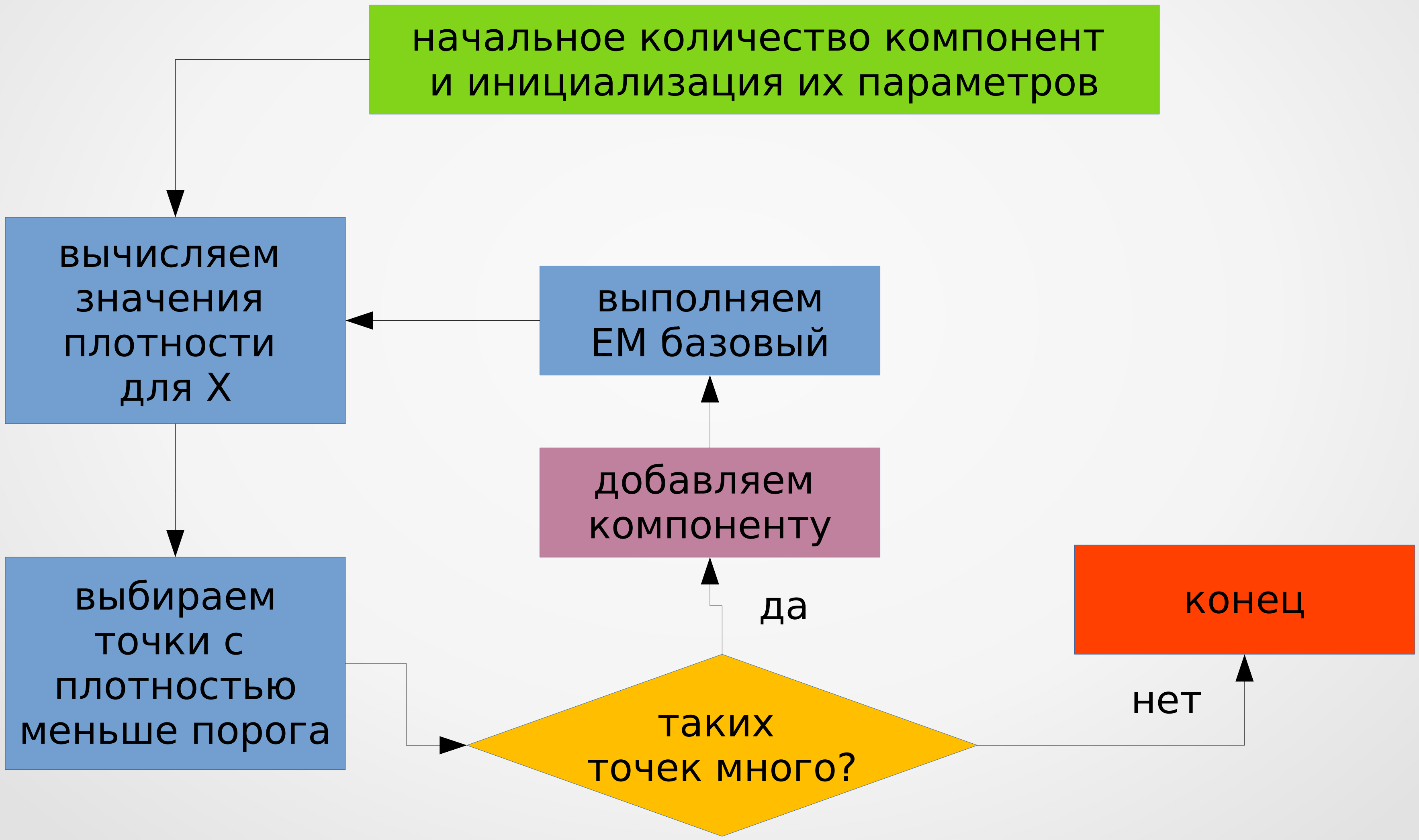
мат.ожидание компоненты

$$\Sigma_j=\frac{1}{c N_j}\sum_{i=1}^{m} g_{ij}(x_i-\mu_j)^T(x_i-\mu_j);0<c\leq 1$$

матрица ковариаций компоненты

Восстановление плотности

ЕМ с последовательным добавлением компонент



Восстановление плотности

EM последовательное добавление компонент (для нормальной плотности)

начальные значения параметров первой компоненты смеси

$$w_1 = 1$$

вес компоненты

$$\Sigma_1 = cov(X)$$

матрица ковариаций компоненты

$$\mu_1 = \frac{1}{|X|} \sum X$$

мат.ожидание компоненты

Восстановление плотности

EM последовательное добавление компонент
(для нормальной плотности)

$X_{low} \subset X$ - точки с правдоподобием (значением смеси) ниже порога

начальные значения параметров новой компоненты смеси

$$w_{k+1} = \frac{|X_{low}|}{|X|}$$

вес компоненты

$$\Sigma_{k+1} = cov(X_{low})$$

матрица ковариаций компоненты

$$\mu_{k+1} = w_{k+1} \frac{1}{|X_{low}|} \sum X_{low}$$

мат.ожидание компоненты

Восстановление плотности

EM последовательное добавление компонент (для нормальной плотности)

$X_{low} \subset X$ - точки с правдоподобием (значением смеси) ниже порога

начальные значения параметров новой компоненты смеси

$$w_{k+1} = \frac{|X_{low}|}{|X|}$$

вес компоненты

$$\Sigma_{k+1} = cov(X_{low})$$

матрица ковариаций компоненты

$$\mu_{k+1} = w_{k+1} \frac{1}{|X_{low}|} \sum X_{low}$$

мат.ожидание компоненты

коррекция весов старых компонент смеси

$$w_i := w_i (1 - w_{k+1})$$

после определения новых параметров смеси запускаем EM

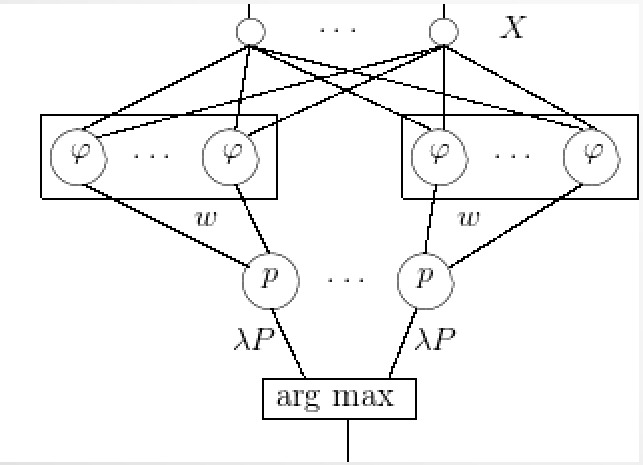
Восстановление плотности

RBF - сеть радиальных базисных функций

Байесовский классификатор

плотности классов - смеси нормальных распределений

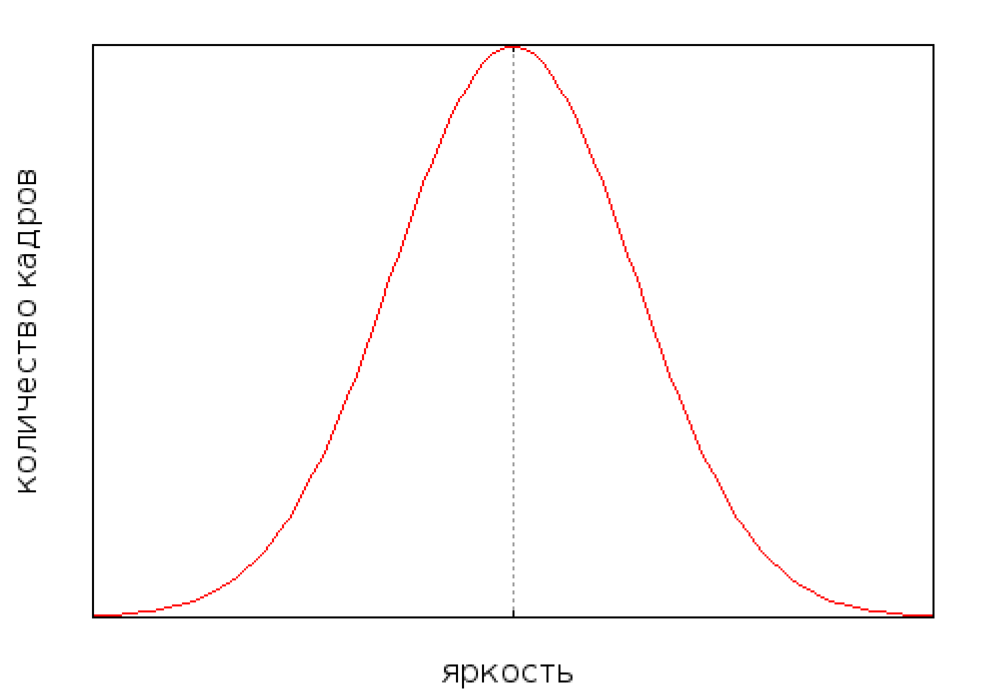
$$a(x) = \underset{y \in Y}{\operatorname{argmax}} \lambda_y P(y) p(x|y)$$



$$p(x|y) = \sum_k w_k^y \varphi_k^y(x; \theta_k^y) = \sum_k w_k^y N(x; \Sigma_k^y, \mu_k^y)$$

Восстановление плотности

Пример: детектор новых объектов для неподвижных камер



Восстановление плотности: литература

git clone *https://github.com/mechanoid5/ml_lectorium.git*

- К.В. Воронцов   Байесовская теория классификации и методы восстановления плотности. - Курс "Машинное обучение" ШАД Яндекс 2014

- Борисов Е.С. Байесовский классификатор. *http://mechanoid.su/ml-bayes.html*

- Борисов Е.С. Восстановление смеси плотностей распределений с помощью EM-алгоритма. *http://mechanoid.su/ml-em-base.html*

- Борисов Е.С. Классификатор на основе RBF. *http://mechanoid.su/ml-rbf.html*

- Борисов Е.С. Детектор объектов для неподвижных камер. *http://mechanoid.su/cv-backgr.html*

Восстановление плотности

Вопросы ?