

# Байесова теория классификации ВТС

Машинное обучение

2022, Корлякова Мария Олеговна

# План

---

- ▶ Формулировка идеи
- ▶ Байессов классификатор
- ▶ EM - алгоритм



# ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 1

---

- ▶ Метод обучения (learning algorithm) это отображение  $\mu: (X \times Y)^\ell \rightarrow A$ , которое произвольной конечной выборке  $X^\ell = (x_i, y_i)_{i=1}^\ell$  ставит в соответствие некоторый алгоритм  $a \in A$ .
- ▶ Метод  $\mu$  строит алгоритм  $a$  по выборке  $X^\ell$ .
- ▶ Метод обучения должен допускать эффективную программную реализацию.



# ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 1

---

- ▶ На этапе обучения метод  $\mu$  по выборке  $X^\ell$  строит алгоритм  $a = \mu(X^\ell)$ .
- ▶ На этапе применения алгоритм  $a$  для новых объектов  $x$  выдаёт ответы  $y = a(x)$ .



# ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 1

---

- ▶ Функция потерь (loss function) :
- ▶  $L(a, x) \geq 0$ ,
- ▶ Характеризует величину ошибки алгоритма  $a$  на объекте  $x$ .
- ▶  $L(a, x) = 0$ , тогда  $y = a(x)$  корректным.
- ▶ Функционал качества алгоритма  $a$  на выборке  $X^\ell$ :
- ▶  $Q(a, X^\ell) = 1/\ell \sum L(a, x)$



# ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 1

---

- ▶ минимизация эмпирического риска
- ▶ Empirical risk minimization, ERM
  
- ▶  $\mu(X^\ell) = \arg \min Q(a, X^\ell)$



## ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 2

---

- ▶ Вероятностная постановка задачи
- ▶ Нет целевой зависимости  $y(x)$
- ▶ вероятностное распределения на множестве  $X \times Y$  с плотностью  $p(x, y)$ ,
- ▶ случайно и независимо выбираются  $\ell$  наблюдений
- ▶  $X^\ell = (x_i, y_i)^\ell$



## ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 2

---

- ▶ Вероятностная постановка задачи

$$X^\ell = (x_i, y_i)^\ell$$

- ▶  $\phi(x, y, \theta)$  - совместная плотность распределения объектов и ответов
- ▶ неизвестная плотность  $p(x, y)$ .
- ▶ ЗАДАЧА: определить значение параметра  $\theta$ , при котором выборка данных  $X^\ell$  максимально правдоподобна,





## ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ - 2

---

- ▶ Вероятностная постановка задачи

Принцип максимума правдоподобия

- ▶  $p(X^\ell) = p((x_1, y_1), \dots, (x_\ell, y_\ell)) = p(x_1, y_1) \dots p(x_\ell, y_\ell)$

- ▶  $p(x_1, y_1) \rightarrow \phi(x_i, y_i, \theta)$

- ▶  $L(\theta, X^\ell) = \prod \phi(x_i, y_i, \theta) \rightarrow \max \theta$

- ▶  $-\ln L(\theta, X^\ell) = -\sum \ln \phi(x_i, y_i, \theta) \rightarrow \min \theta$



# Вероятностные

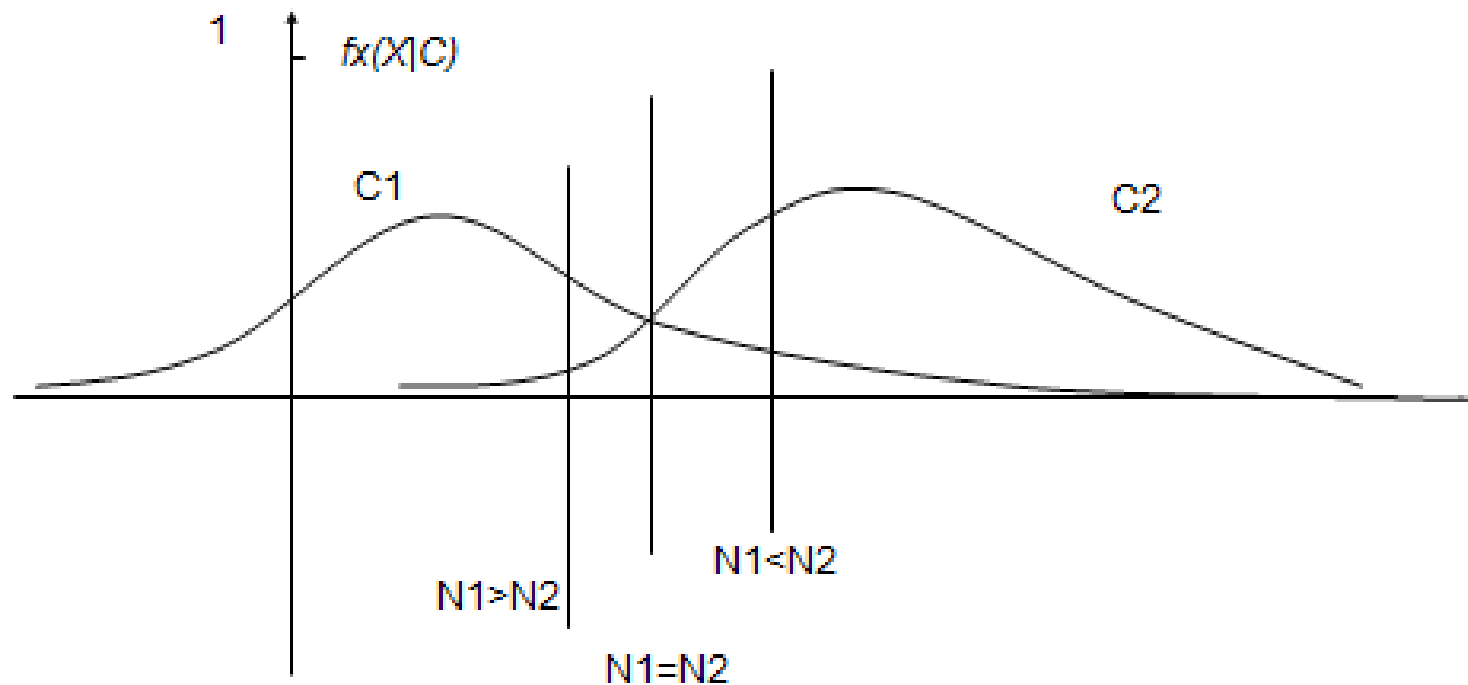
---

- ▶ Построение классификации на известных распределениях
  - ▶ Область: Задачи с известным распределением.
- ▶ Оценках плотностей распределения значений признаков (или сходства и различия объектов)
- ▶ **Проблема:** Необходимость набора большой статистики. Необходимость перебора всей обучающей выборки при распознавании, высокая чувствительность к непредставительности обучающей выборки и артефактам



# Вероятностная картина классификации

- ▶ Плотностей распределения классов вдоль значений признаков



# Вероятностный классификатор

---

- ▶ Построение аппроксимации гауссовой смеси многомерного распределения
- ▶ Оценивание плотности условной вероятности по выборке  $T = \{X, y\}^I$
- ▶ Частотная оценка априорной вероятности:  
 $P(y=k) = |T(y=k)|/|T|$
- ▶ Параметрическая оценка плотности:  
 $p(X|y) = f(X, q_y)$ ,  $q_y = \arg \max \sum (\log f(X_i, q_y)) \quad X_i \in T(y = i)$
- ▶ Непараметрическая оценка плотности:  
 $p(X|y) = \sum (w_y P(y)/|T_y|) \sum K(r(X, x_i)/h)$  – Parzen Window
- ▶ GaussianMixtureModel



# Байесов классификатор

---

$X$  — объекты,  $Y$  — классы,  $p(x,y)$  - плотность

►  $p(x,y) = p(x)P(y|x) = P(y)p(x|y)$

-  $P(y)$  — априорная вероятность класса  $y$

-  $p(x|y)$  — функция правдоподобия

-  $P(y|x)$  — апостериорная вероятность класса  $y$

$$P(y|x) = P(y)p(x|y)/p(x)$$

► Байесов классификатор

$$a(x) = \arg \max P(y|x) = \arg \max P(y) p(X|y)$$

# Байесов классификатор

---

- ▶ **Теорема.** Байесовский классификатор является оптимальным по отношению к минимизации вероятности ошибки классификации.



# Вероятностный классификатор

---

- ▶ Построение аппроксимации гауссовой смеси многомерного распределения
- ▶ Оценивание плотности условной вероятности по выборке  $X^l$
- ▶ Переменные  $X=(x_1, x_2, \dots, x_n)$  – независимы (очень опасно)
- ▶  $p(X|y) = p(x_1|y) p(x_2|y) \dots p(x_n|y) \rightarrow \ln(P(X|y))$
- ▶  $a(x) = \arg \max_y w_y P(y) P(X|y)$
- ▶  $a(x) = \arg \max (\ln(w_y P(y)) + \ln(p(X|y)))$
- ▶  $a(x) = \arg \max (\ln(w_y P(y)) + \sum \ln(p(x_i|y)))$



# Вероятностный классификатор

---

Пусть  $p(x_i|y)$  – экспоненциальны ( $q_i = \langle M_i, D_i \rangle$ )

►  $a(x) = \arg \max (\ln(w_y P(y)) + \sum \ln(p(x_i|y)))$

$\max(L = \sum \sum (\sum \ln(p(x_i|y; q_i)))$  по всем примерам и всем классам

Для каждого каждого класса отдельно:

$\max(L_i = (\sum \ln(p(x_i|y; q_i))))$

Аналитически не всегда решается:

- $M_i, D_i$  – среднее класса по признакам, СКО класса по признакам

---

► Тогда классификатор линейный



# Байесов классификатор

---

- ▶  $f_X(X/C1) = 1/(2*\pi * D_1^2) * \exp(-1/(2*D_1^2)*||X-M_1||^2)$
- ▶  $f_X(X/C2) = 1/(2*\pi * D_2^2) * \exp(-1/(2*D_2^2)*||X-M_2||^2),$

▶ где

$f_X(X/C1)$ - функция плотности условной вероятности для класса C1,

$f_X(X/C2)$ - функция плотности условной вероятности для класса C2,

$D_1^2$  – дисперсия класса C1,  $M_1$  – вектор средних значений по всем признакам класса C1,  $||.||$  - оператор вычисления расстояния по Евклиду,

$D_2^2$  – дисперсия класса C2,  $M_2$  – вектор средних значений по всем признакам класса C2.



# Байесов классификатор

---

- ▶ отношение правдоподобия

$\Lambda(X) > \xi$  для класса  $C1$ , где

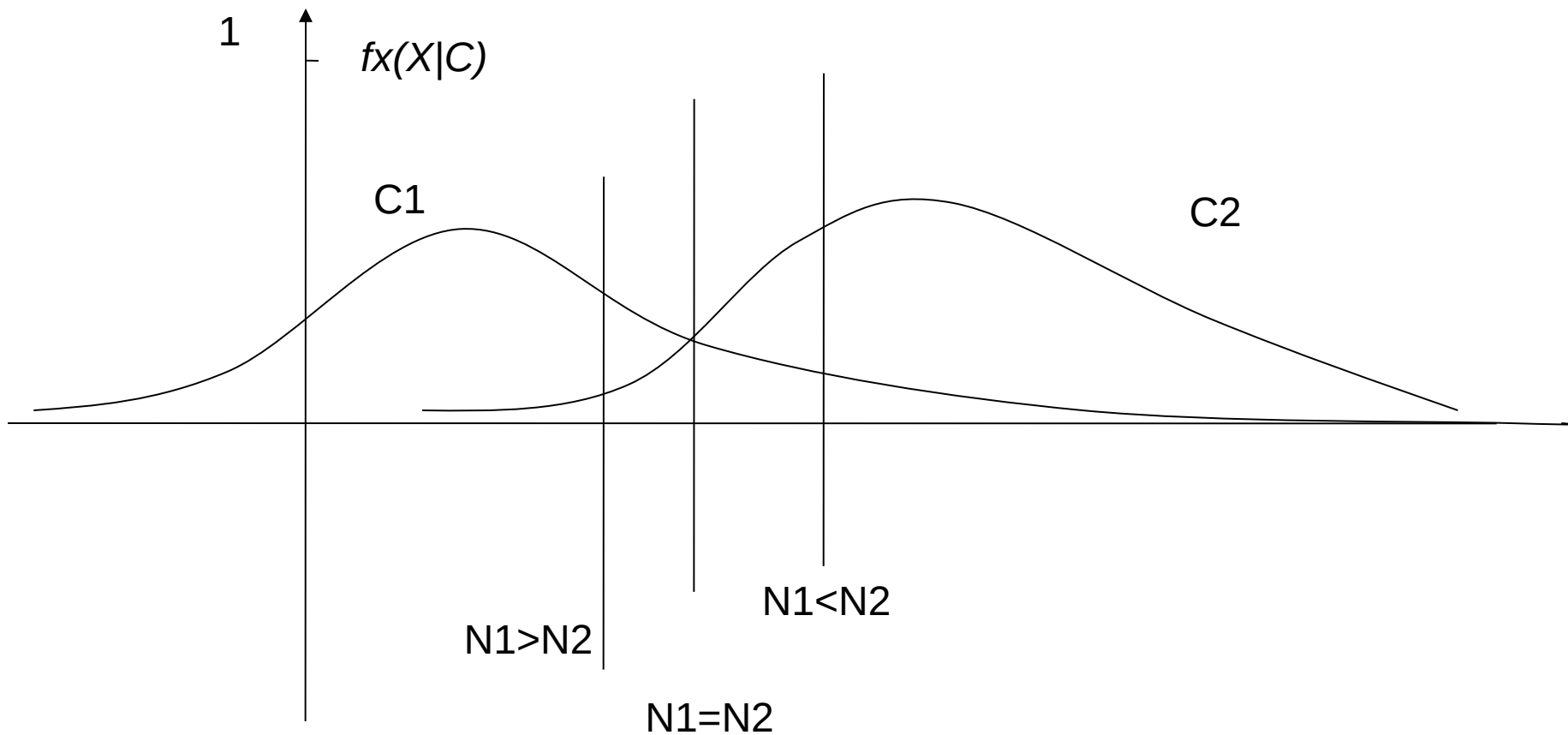
$$\Lambda(X) = f_X(X/C1)/f_X(X/C2), \quad \xi = p1/p2,$$

$p_i$  – априорная вероятность класса  $C_i$ ,



# Байесов классификатор

---



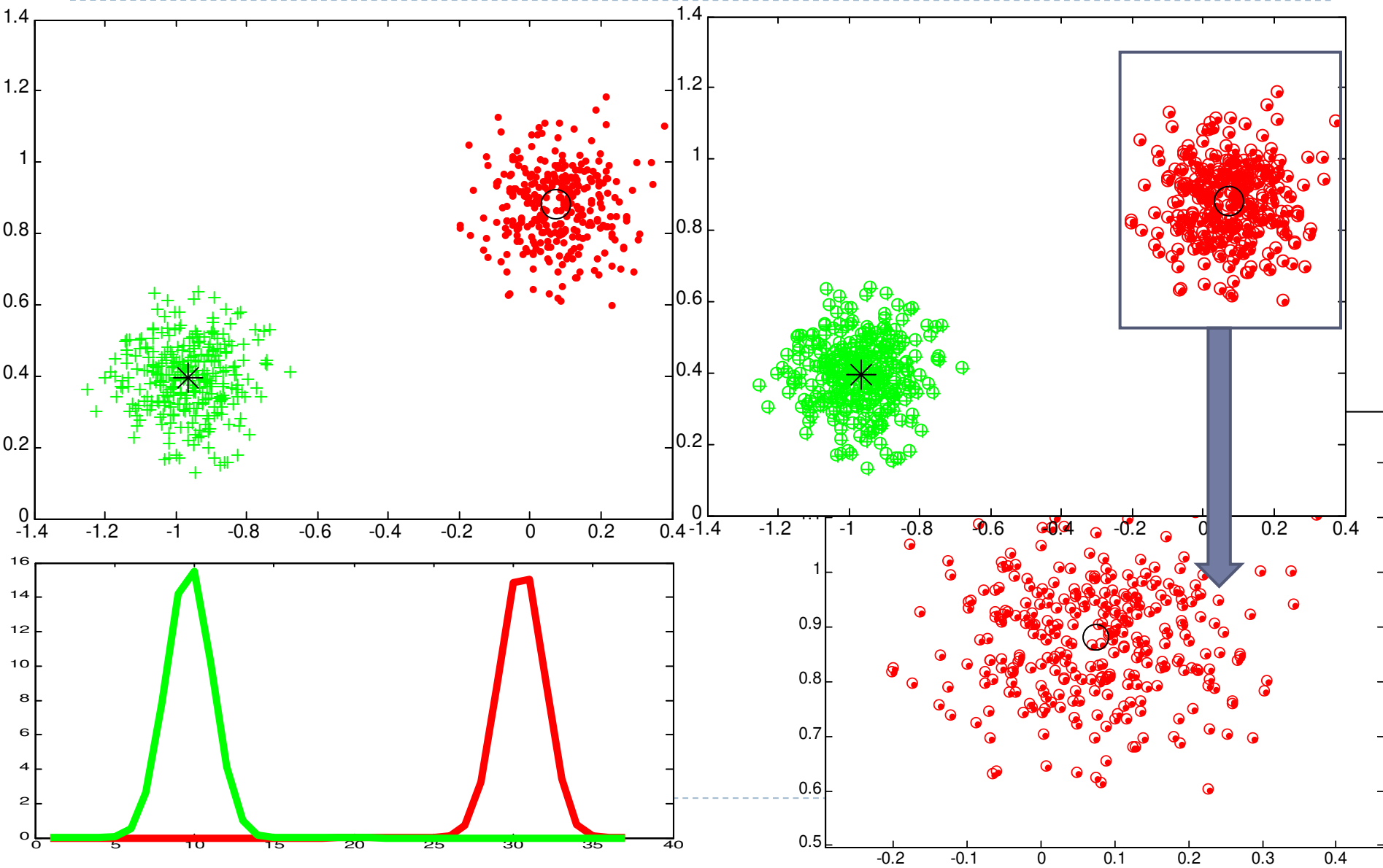
# Байесов классификатор

---

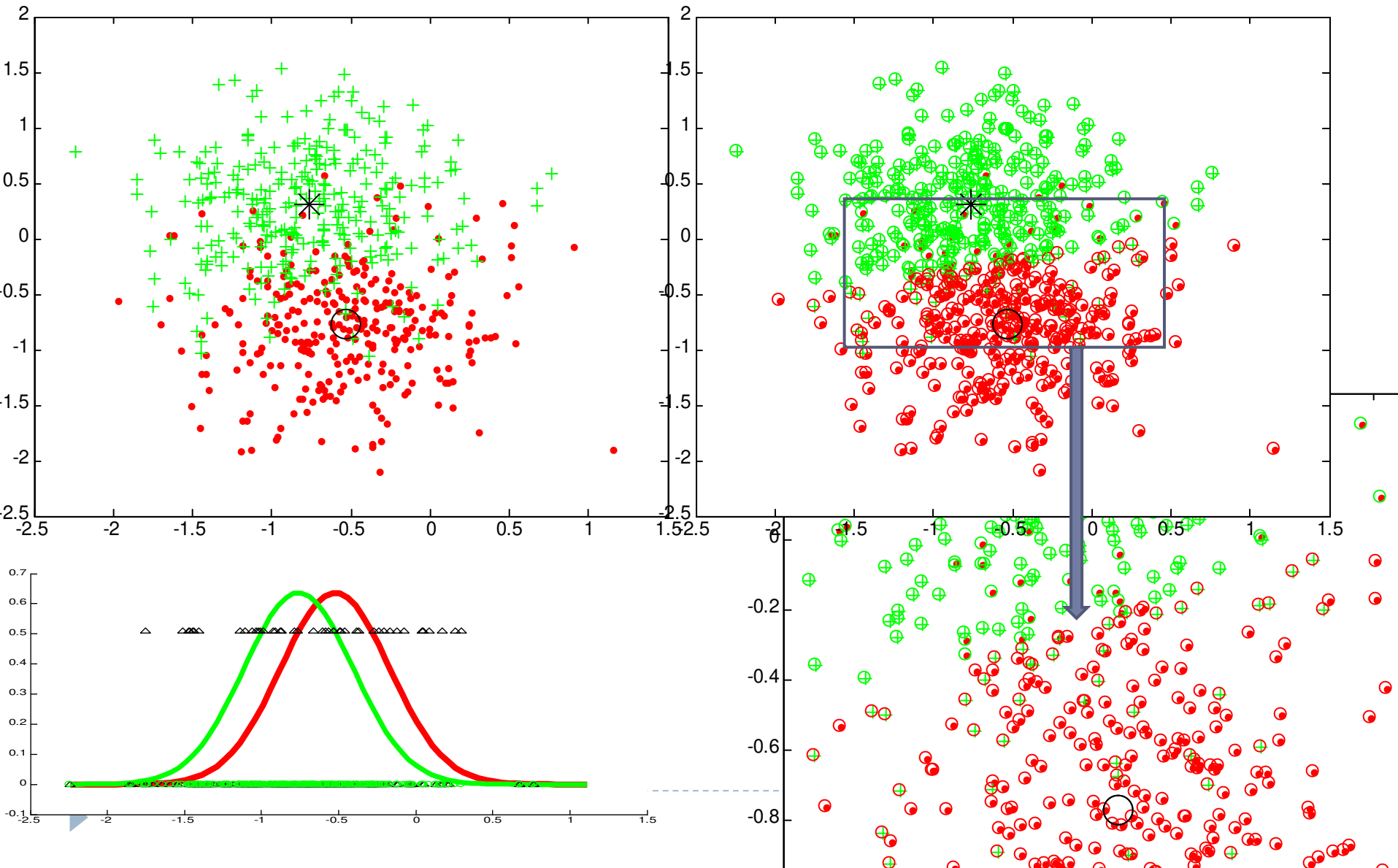
- ▶ вероятность ошибки классификатора
- ▶  $P_e = p_1 * P(e|C_1) + p_2 * P(e|C_2)$ , где
- ▶  $P(e|C_i)$  – условная вероятность ошибки для входного вектора класса  $i$  (установлена по фактическому отнесению примера к классу  $i$  байесовым классификатором),
- ▶  $e$  - множество результатов некорректной классификации



# Байесов классификатор: $D=0.1$



# Байесов классификатор: $D=0.5$



# Вероятностный классификатор

---

- ▶ Построение аппроксимации гауссовой смеси многомерного распределения
- ▶ Оценивание плотности условной вероятности по выборке  $T = \{X, y\}^I$

- ▶ Непараметрическая оценка плотности:

$$p(X|y) = \sum (w_y P(y)/|T_y|) \sum K(r(X, x_i)/h) - \text{Parzen Window}$$



# Вероятностный классификатор

---

- ▶ Построение аппроксимации гауссовой смеси многомерного распределения
- ▶ Оценивание плотности условной вероятности по выборке  $T = \{X, y\}^I$
- ▶ GaussianMixtureModel





# ЕМ - алгоритм

---

- ▶  $P(x) = \sum (w_j f(x_i, q_j))$  - порождающая модель смеси
  - ▶  $\sum w_j = 1$
  - ▶  $L(w, q) = \sum_{i=1, m} \ln \sum_{j=1, k} (w_j f(x_i, q_j)) \rightarrow \max (q, w)$ 
    - $i=1, m \quad j=1, k$
  - ▶ k- ядер
  - ▶ **Expectation**
  - ▶  $g_{ij} = w_j f(x_i, q_j) / \sum_{j=1, k} (w_j f(x_i, q_j)) \quad j = 1, k$ 
    - $i=1, m \quad j=1, k$
  - ▶ **Maximization**
- $q_j = \arg \max \sum_{i=1, m} g_{ij} \ln(f(x_i, q_j)), \quad w_j = 1/m \sum_{i=1, m} g_{ij}$



# Литература

---

- ▶ Саймон Хайкин. Нейронные сети полный курс. – М.: ООО «И.Д. Вильямс», 2006.
- ▶ Математические методы распознавания образов. Курс лекций. МГУ, ВМиК, кафедра «Математические методы прогнозирования», Местецкий Л.М., 2002–2004.
- ▶ Машинное обучение (ШАД) — Яндекс
- ▶ Лекции Воронцов К.

