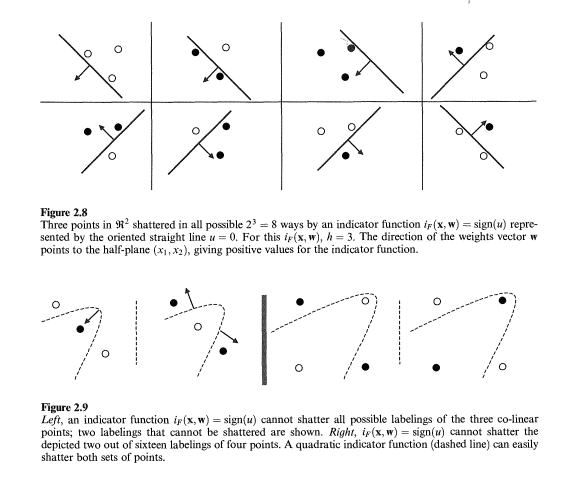
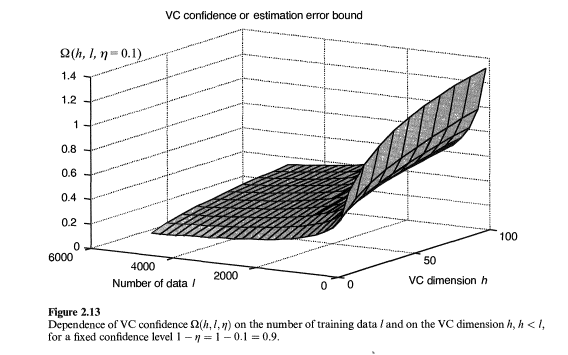
*Пример верхней границы ожидаемого риска для задачи бинарной классификации*.

  (1) Класс функций {*f(****x****,*α*)*} имеет меру сложности *h*.

Размерность Вапника-Червоненкиса(**VC-dimension**) (*h*) для множества функций {*f(****x****,*α*)*}. . 



**Линейный SVM-классификатор**

*Линейно разделимые данные.*

Рассмотрим на примере задачи бинарной классификации линейно разделимых данных основные идеи и преимущества SVM-обучения.

Пусть (***x****1*,*y1*), … , (***x****l*, *yl*) ∈ *Rn*×*R* − тренировочная последовательность (ТП).

Выход *y*  : −1 и +1 (метка класса) .

*f*(***x***)=0, где *f*(***x***)=(***w***, ***x***)+*b*,  ***x*** , ***w*** ∈ *Rn* ; *b* ∈ *R* .

Функция потерь: *θ*(–*yf*(*x*)), где *θ*(⋅) – функция Хевисайда.

**Каноническая** гиперплоскость **для данной ТП** :.

2/||***w***|| – **margin**, **ширина полосы**,  – евклидова норма.

**Поддерживающие** гиперплоскости : (***w***, ***x***) + *b* = ±1 .



Рис.1. Каноническая разделяющая гиперплоскость и соответствующие ей поддерживающие гиперплоскости.

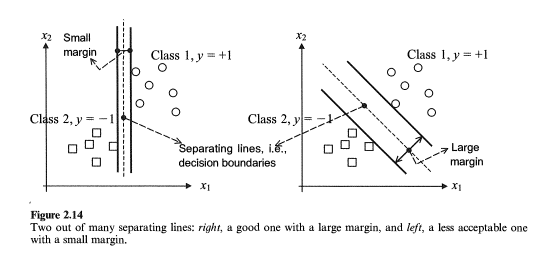


Рис.2. Две возможные разделяющие гиперплоскости с разной шириной полосы. Решение, представленное справа, предпочтительнее, т.к. обладает большей способностью к обобщению

Если 2/||***w***|| >= 2/*A*, то ||***w***|| ≤ *A* и *h* ≤ *A2R2+*1,

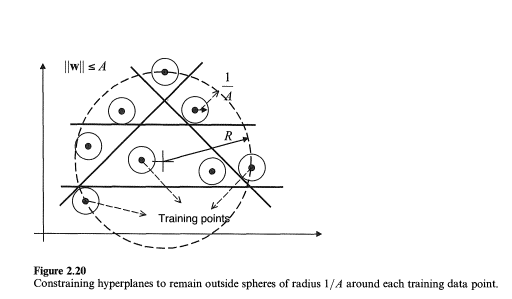
где *R* – радиус наименьшей гиперсферы, охватывающей множество векторов ТП. 

Рис.3. Наименьшая гиперсфера, охватывающая множество векторов ТП. Разделяющие гиперплоскости остаются за пределами гиперсфер радиуса 1/*А*, охватывающих точки ТП.

Малое значение ||***w***|| приводит к малому *h* (и малому Ω(*h*)), и **минимизация** ||***w***|| представляет собой **реализацию принципа структурной минимизации риска**.



 (2)

α*i* ≥ 0 − множители Лагранжа.

 (3)

 (4)

 (5)

Условия Каруша-Куна-Таккера (ККТ):

 (6)

 (7)

 (8)

 (9)

 (10)

Решение задачи бинарной классификации линейно разделимых данных представляет собой оптимальную гиперплоскость :

 (11)

= – **линейная машина опорных векторов** (**hard margin SVM**).

Для тестирования (предсказания класса) новых образцов используют:

 (12)

*Линейно неразделимые данные.*

Построим **soft margin SVM**.

Введем неотрицательные переменные ξ*i* , *i*=1,…,*l*. (slack variables)



*C* > 0 – **параметр регуляризации**.

, (13)

где, μ*i* – множитель Лагранжа для ограничений на переменные ξ*i*, *i*=1, …, *l*.



Условия оптимальности (условия ККТ):

 (14)

 (15)

 (16)

 (17)

 (18)

 (19)

 (20)

  (21)

 (22)

Условия (14)–(22) для **оптимальных** множителей Лагранжа в более удобной форме:

* если α*i* = 0, то *yi f*(***x****i*) ≥ 1 и ξ*i* = 0;
* если 0 < α*i* < *C*, то *yi f*(***x****i*) = 1 и ξ*i* = 0; (соответствующие – margin(al)SVs);
* если α*i* = *C*, то *yi f*(***x****i*) ≤ 1 и ξ*i* ; (соответствующие – bounded SVs).

SVM-классификатор для линейно неразделимых данных

 по-прежнему является **разреженным по α**.



Рис. 4. На поддерживающих гиперплоскостях лежат опорные векторы (обведены кружками), соответствующие им α : 0 < α < C. Внутрь полосы могут попасть связанные опорные векторы, соответствующие им α = С. Для правильно классифицированных векторов, лежащих вне полосы, α = 0 .

