**Ядерные методы**

План занятия

1*. История вопроса*.

Aronszajn (1950); Parzen (1962); Aizerman et al (1964).

Процедуру, преобразующую линейный алгоритм в более общий ядерный метод, называют "ядролизацией" (kernelization).

Ядра могут использоваться для обобщений тех алгоритмов, которые формулируются в терминах скалярных произведений. В последние годы наблюдается большое количество "ядролизаций" алгоритмов для различных задач (среди них: SVM, PCA (principal component analysis), линейный дискриминантный анализ).

2. *Представление данных*.

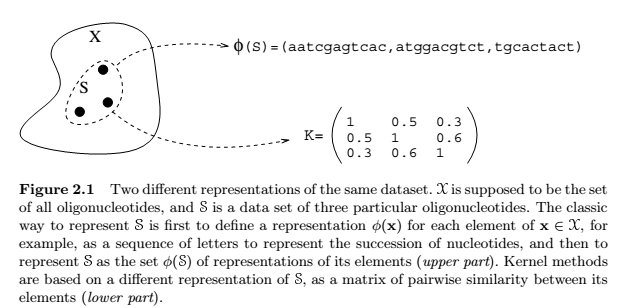
, ... ,. Как представить данные для дальнейшей обработки?

 , ... ,.

При использовании ядерных методов данные **никогда** не представляются индивидуально а только через множество попарных сравнений. Т.е. вместо некоторого отображения из исходного пространства объектов в пространство признаков используется вещественнозначная "функция сравнений" (ядерная функция). И множество из *l* объектов, подлежащих анализу, представляется *l×l* **матрицей попарных сравнений**.

*k* : *X* × *X* → *R* , ***K*** ={****}, *i*, *j* =1,...,*l* .

Все ядерные методы предназначены для обработки такой **ядерной матрицы**.



3. *Положительно определённые ядра*.

Функция *k* : *X* × *X* → *R* – **положительно определённое ядро**, т. и т.т. , когда она симметрична по своим аргументам и

, набора , ... ,; набора ,...,  , 



Если *k*(***x***, ***x***′) – **положительно определённое ядро**, то любая матрица попарных сравнений, построенная с его использованием (матрица Грама), симметрична и **положительно полу-определена** (неотрицательно определена).

4. *Ядро как скалярное произведение*.

Показано, что класс ядер вида *k*(***x***, ***x***′)=(***Φ***(***x***), ***Φ***(***x***′)) совпадает с классом положительно определённых ядер.

Любое положительно определённое ядро *k* можно интерпретировать как скалярное произведение в некотором пространстве (пространстве признаков). При этом представление любого объекта ***x*,** *Φ*(***x***)**,** не нужно вычислять явно, – только попарные скалярные произведения необходимы.

Поскольку

,

ядра вида  *k*(***x***, ***x***′) = (***Φ***(***x***), ***Φ***(***x***′)) положительно определены при любом выборе отображения ***Φ***.

5. *Ядро как мера подобия объектов*.

Интерпретация ядра как меры подобия объектов полезна при конструировании ядер для различных типов данных (например, строк, графов, деревьев и пр.). Использование подходящих ядерных функций позволяет применять SVM-методы и при анализе данных, не имеющих ясного векторного представления.

Например, гауссово ядро – убывающая функция евклидова расстояния между двумя векторами,



имеет подходящую интерпретацию как мера подобия: чем больше *k*(***x***, ***x***′), тем ближе ***x*** и ***x***′ в пространстве .

6. *Ядро как мера гладкости функций*.

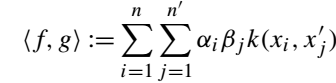
Понятие " гладкость функций " двойственно понятию " подобие объектов ". Функция "гладкая", если она мало изменяется между "подобными" точками.

Пусть ***Φ***(***x***) = *k*(***x***, ∙) – функция, которая назначает значение *k*(***x***, ***x***′) элементу ***x***′ , *k* : *X*×*X* → *R* – положительно определённое ядро.

Построим пространство со скалярным произведением, содержащее образы исходных объектов при отображении ***Φ***.

Для этого рассмотрим пространство с элементами вида *f*(∙) = .

*g*(∙) = .

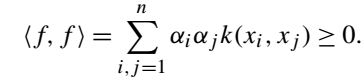
 (++)

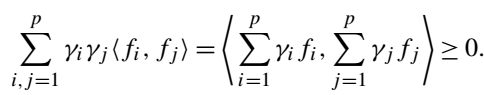
Покажем, что это скалярное произведение.

Заметим, что

 ,  .

;

*f* 

 ,  

Следовательно,  – положительно определённое ядро.

Заметим, что из (++) следует :

, в частности, .

Неравенство Коши-Шварца:

*f* :

Отсюда, если 〈 *f*, *f* 〉 = 0, то *f*  = 0.

Пусть *H* = {*f*} – класс функций *f* : *X* → *R* : *f*(∙) =

Показали, что (++) – скалярное произведение в пространстве *H* ,

а *k* : *X*×*X* → *R* – **репродуктивное ядро** для *H*.

(++) определяет норму в *Н* : 

Таким образом, *H* – гильбертово пространство с репродуктивным ядром (RKHS).

Теорема (Moore-Aronszajn) устанавливает, что для любого положительно определённого ядра RKH-пространство и обратно.

Норма в пространстве *H* является мерой гладкости или мерой сложности функций.

Многие ядерные методы (и SVM в том числе) можно понимать как алгоритмы, которые для данного множества объектов восстанавливают функцию из пространства *Н* , минимизирующую функционал, – регуляризованный риск:

.

Нормаобеспечивает подходящую гладкость решения.

Параметр регуляризации λ > 0.

7. *Теорема репрезентативности* (*representer theorem*).

**Теорема** . 

Пусть ***F*** = {*f*} , *f* : *X* → *R* – RKH-пространство, *k* : *X*×*X* → *R* – репродуктивное ядро для ***F***.

Тренировочная последовательность (***x****1*,*y1*), … , (***x****l*, *yl*) ∈ *X*×*R*.

,  **:** матрица {}имеет максимальный ранг.

Тогда:





[Функция, минимизирующая регуляризованный риск, принадлежит линейной оболочке множества (...). ]

Различные формулировки теоремы см. : [1] – часть 1, гл.2, п. 2.3.3, стр. 47–48;

[2] – п. 2.3.1, стр 1185-1186 ;

[3] – стр. 47–48

Функция, минимизирующая регуляризованный риск, должна иметь вид

,

где вектор весов признаков **β**∈*R l* .

Различные классы функций могут быть обучены путем использования различных ядер.

8. *Ядерный приём*.

Любой линейный алгоритм, использующий только скалярные произведения, может быть неявно выполнен в пространстве признаков, надо лишь каждое скалярное произведение заменить нелинейным ядром. **Нелинейность не требует дополнительных вычислительных затрат.**

Литература

1. *Kernel Methods in Computational Biology*. Bernhard Scholkopf, Koji Tsuda, and Jean-Philippe Vert, editors. The MIT Press, 2004.

2. KERNEL METHODS IN MACHINE LEARNING. Thomas Hoffman, Bernhard Scholkopf and Alexander Smola. The Annals of Statistics Vol. 36, No. 3, p. 1171–1220, 2008.

3. A Tutorial on Support Vector Regression. Alexander Smola and Bernhard Scholkopf NeuroCOLT2 Technical Report Series. October,1998.

См. [1] – часть 1, гл.2, стр. 35–70;

[2] – п. 1, п. 2, п. 6;

[3] – стр. 47–48.