

ИСПОЛЬЗОВАНИЕ БИБЛИОТЕКИ SCIKIT-LEARN В ЗАДАЧЕ КЛАССИФИКАЦИИ ДВИЖЕНИЙ ТРЕКБОЛА

Пархоц К.Г. Костюк Д.А.

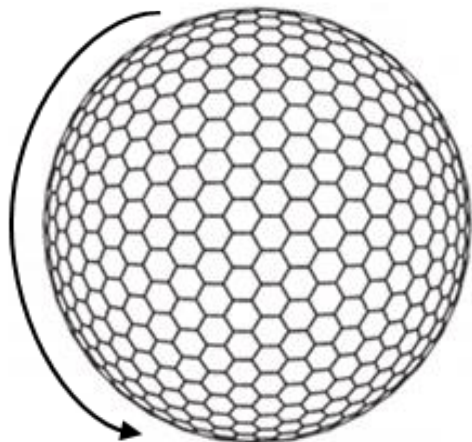
Виды трекболов



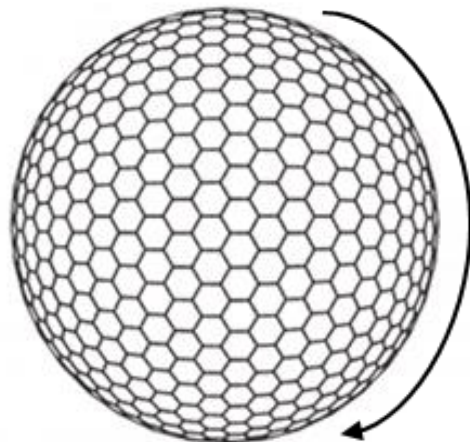
- С помощью кругового вращения шарика трекбола реализована прокрутка в модели kensington slimblade trackball (аппаратными средствами)
- Программная реализация позволила бы добавить подобный функционал в множество более дешёвых трекболов, не оснащённых колесом прокрутки



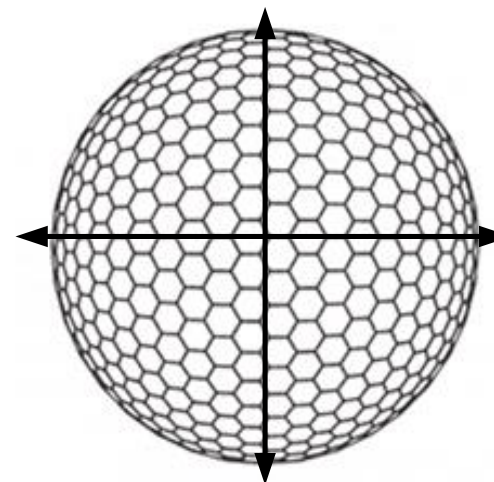
Типы движений



Прокрутка вниз



Прокрутка вверх



Игнорируется

Машинное обучение (англ. Machine Learning)

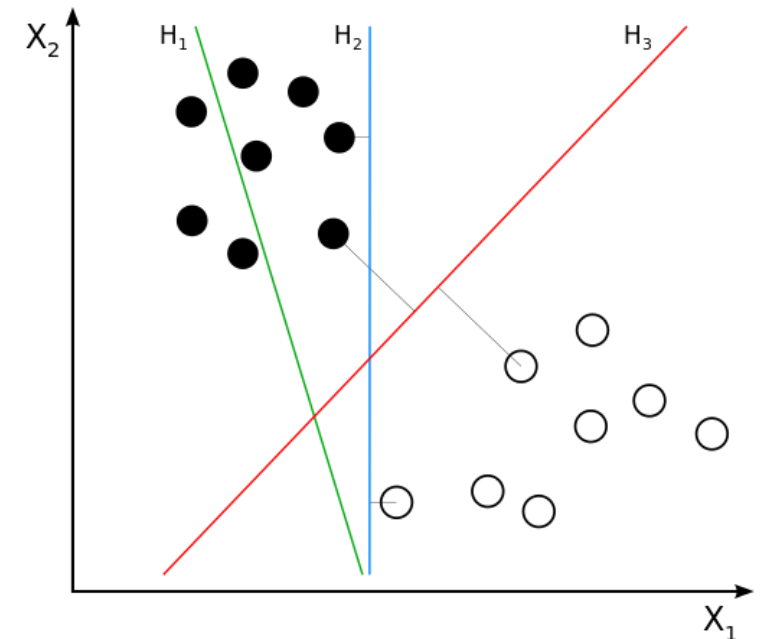
Класс методов искусственного интеллекта

...вместо прямого решения задачи используется обучение в процессе применения к решению множества сходных задач

Могут быть основаны на средствах математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, и др.

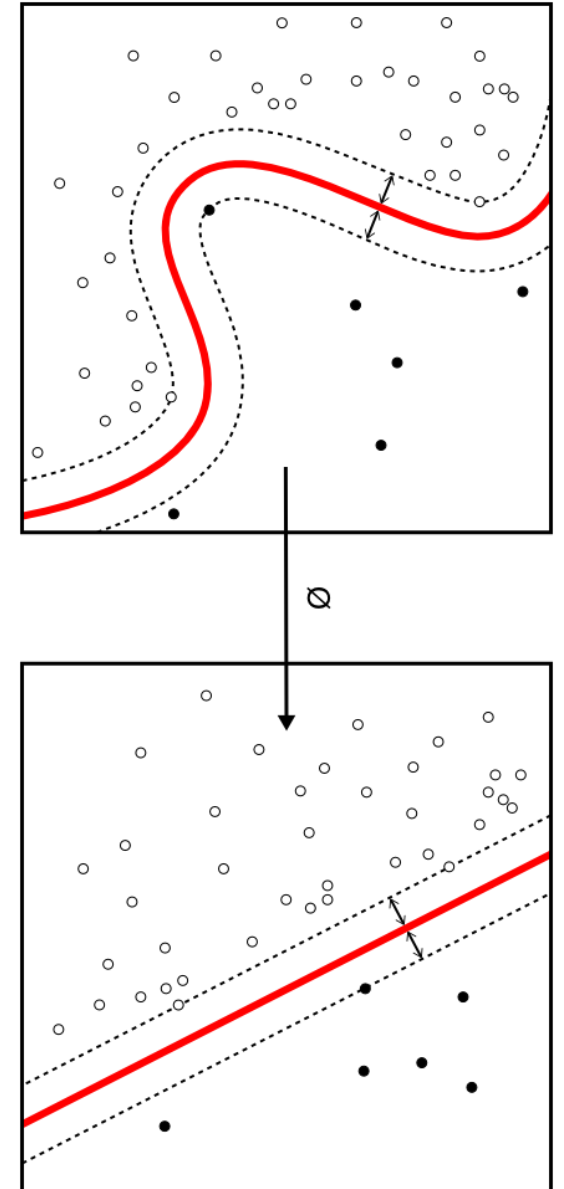
Больше о методе опорных векторов

- Каждый объект данных представляется как вектор (точка) в p -мерном пространстве
 - упорядоченный набор p чисел.
- Каждая из этих точек принадлежит одному из двух классов.
- Требуется разделить точки гиперплоскостью размерности $p-1$.
 - Это — типичный случай линейной делимости.
- Искомых гиперплоскостей может быть много
 - максимизация зазора между классами делает классификацию надёжнее



Svm — нелинейная задача

- Способ создания нелинейного классификатора, известен как «kernel trick»
 - Позволяет строить нелинейные разделители
- Каждое скалярное произведение в формулах построения гиперплоскости заменяется на некую нелинейную функцию
 - Её считают «скалярным произведением в пространстве с ещё большей размерностью»
 - Может использоваться полином, функция гаусса, сигмоидная функция (на основе гиперболического тангенса)



scikit-learn

Это библиотека алгоритмов машинного обучения для языка Python

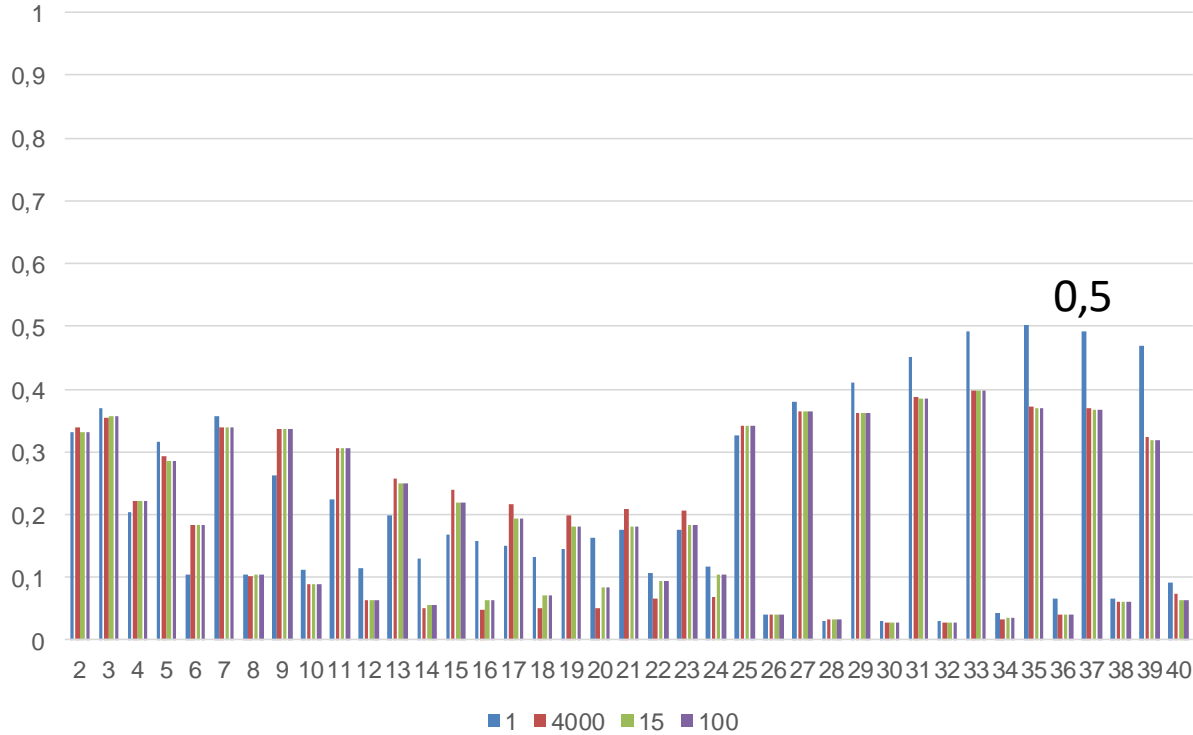
- Включает средства классификации
 - отнесение произвольного объекта из исходного множества к одному из классов, на которые поделено конечное подмножество объектов (выборка)
- регрессионного анализа
 - исследования влияния одной или нескольких независимых переменных на заданную зависимую переменную
- кластерного анализа
 - Упорядочивание объектов в сравнительно однородные группы

Время выполнения операций в секундах для разных библиотек машинного обучения на наборе данных Madelon

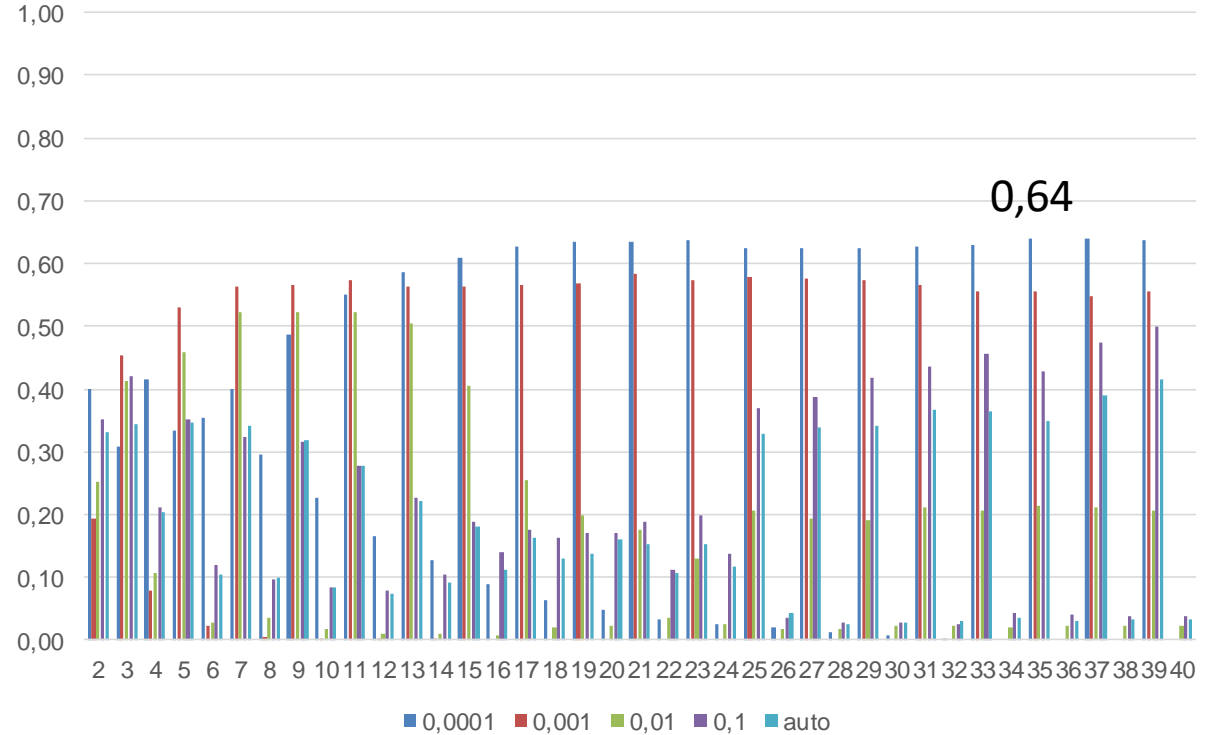
	Scikit-learn	mlpy	pybrain	pymvpa	mdp	Shogun
Support Vector classification	5.2	9.47	17.5	11.52	40.48	5.63
Lasso(LARS)	1.17	105.3	-	37.35	-	-
Elastic Net	0.52	73.7	-	1.44	-	-
K-Nearest Neighbors	0.57	1.41	-	0.56	0.58	1.36
PCA(9 components)	0.18	-	-	8.93	0.47	0.33
k-Means(9 clusters)	1.34	0.79	-	-	35.75	0.68
License	BSD	GPL	BSD	BSD	BSD	GPL

Сигмоидная функция

Коэффициент C

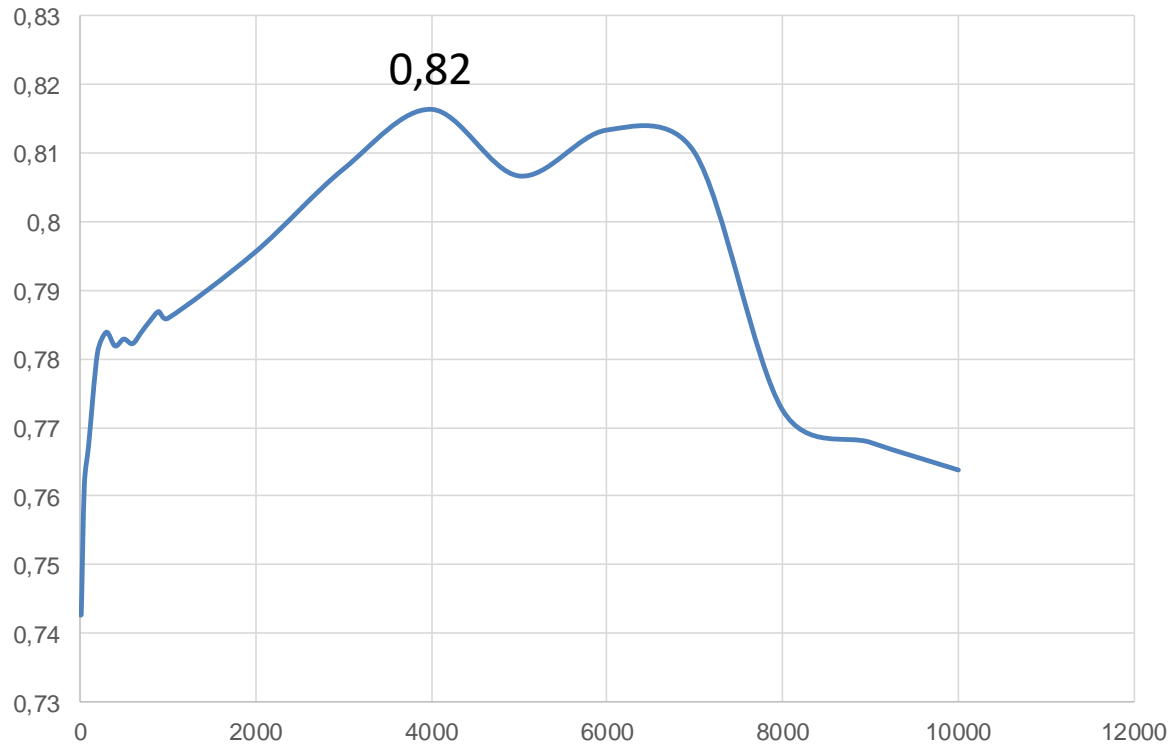


Коэффициент gamma

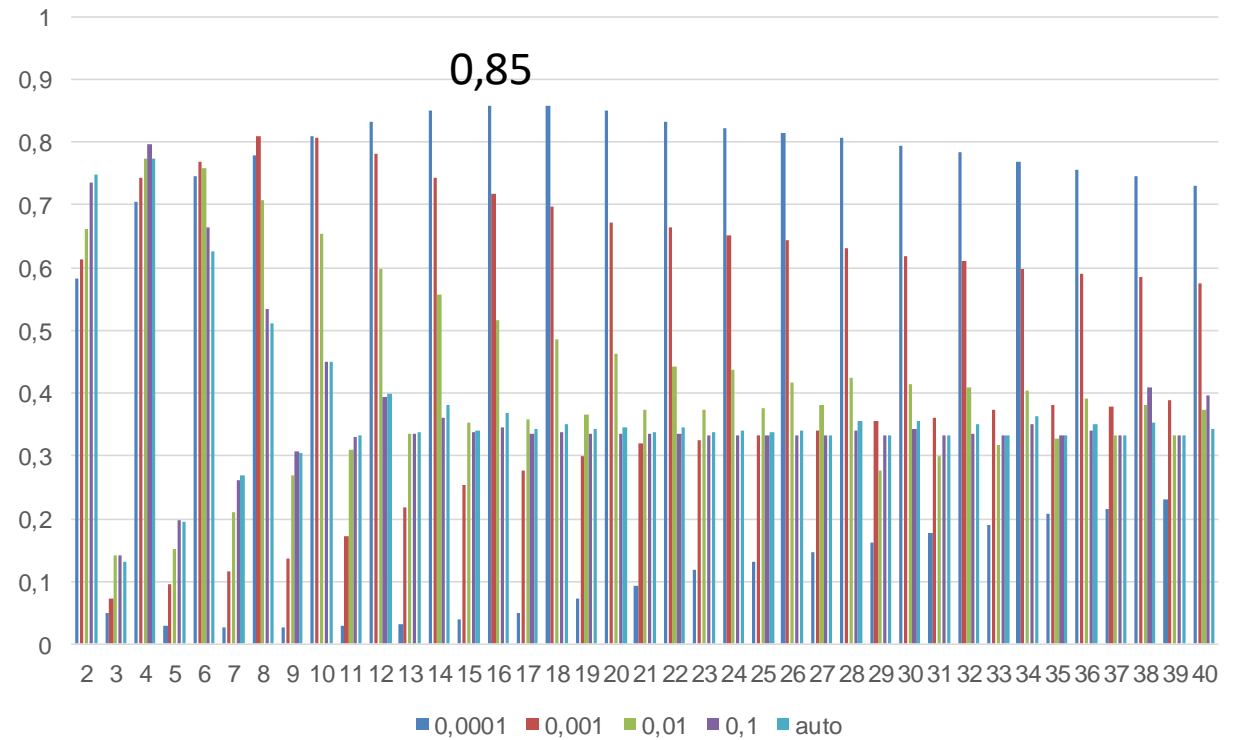


Радиальная базисная функция

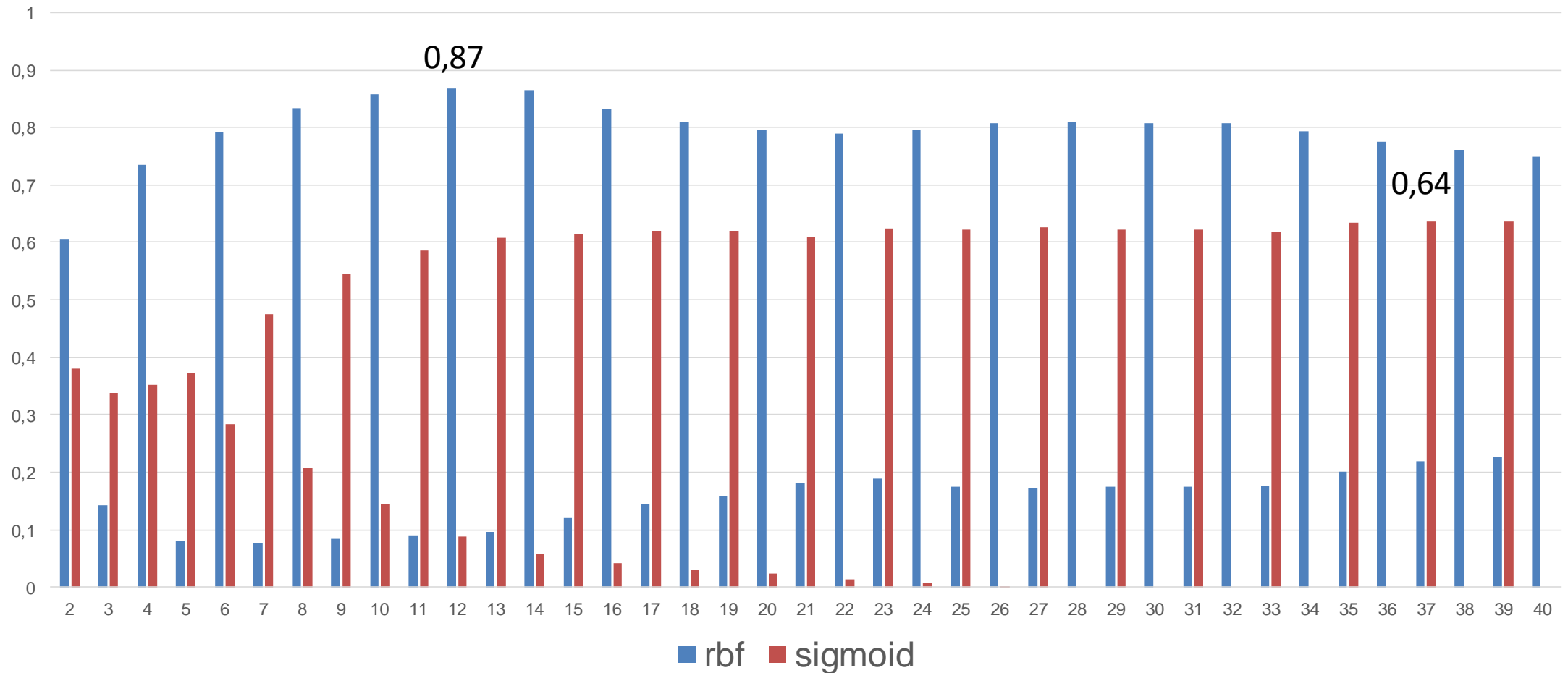
Коэффициент C



Коэффициент gamma



Сравнение ядер



Спасибо за внимание