Министерство образования Республики Беларусь

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет Компьютерных сетей и систем

Кафедра Информатики

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

по курсу «Машинное обучение»

Реализация криптографических атак с помощью машинного обучения на физически неклонируемые функции

Студент:
гр. 758641
Ярош Г.И.

Проверил:
Заливако С. С.

Минск, 2018

СОДЕРЖАНИЕ

РЕАЛИЗАЦИЯ КРИПТОГРАФИЧЕСКИХ АТАК С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ НА ФИЗИЧЕСКИ НЕКЛОНИРУЕМЫЕ ФУНКЦИИ

1. Цель

Создать модель, которая могла бы предсказать ответы ФНФ по запросам, которых нет в обучающей выборке. При построении модели применить различные методы машинного обучения. Исследовать эффективность применения модели при различных входных параметрах.

2. Оценка качества моделей

Решаемая задача является задачей бинарной классификации. Ответы ФНФ 1 и 0 являются равнозначными. Следовательно, для оценки качества можно использовать метрику точности, описываемую по следующей формуле:

$$accuracy(y, \hat{y}) = \frac{1}{L}\sum_{i=0}^{L-1} 1(y_i = \hat{y_i}),$$

где y – значение ответов ФНФ, $\hat{y}$ – предсказанные ответы ФНФ, L - размер выборки.

Значения точности предсказания должны значительно превосходить 0.5, т.к. данный уровень достижим при случайном угадывании ответов ФНФ. Исходная выборка должна быть разделена на тренировочную и тестовую части. Каждая модель должная быть обучена на тренировочной, а затем должна быть подсчитана точность между необходимыми ответами ФНФ и предсказанными моделью для тестовой выборки.

3. Исследование эффективности различных алгоритмов

Для реализации необходимой модели мною были применены следующие алгоритмы машинного обучения: логистическая регрессия, метод опорных векторов, градиентный бустинг.

С помощью каждого алгоритма была обучена модель и проведено сравнение эффективности их применения. Предварительно, все входные запросы были преобразованы в знаковый вид по формуле:

$$c'_l = \prod_{i=0}^{l}(1 - 2c_i), l = 0, \dots, N;$$

Зависимость точности предсказания различных моделей от размера обучающей выборки при длине запроса ФНФ в 64 и 128 бит приведена на рисунках 1 - 6.

Обучение проводилось на выборке длинной от 16 до 10^5 запросов и ответов. Точность подсчитывалась на выборке размером $9 \cdot 10^5$.

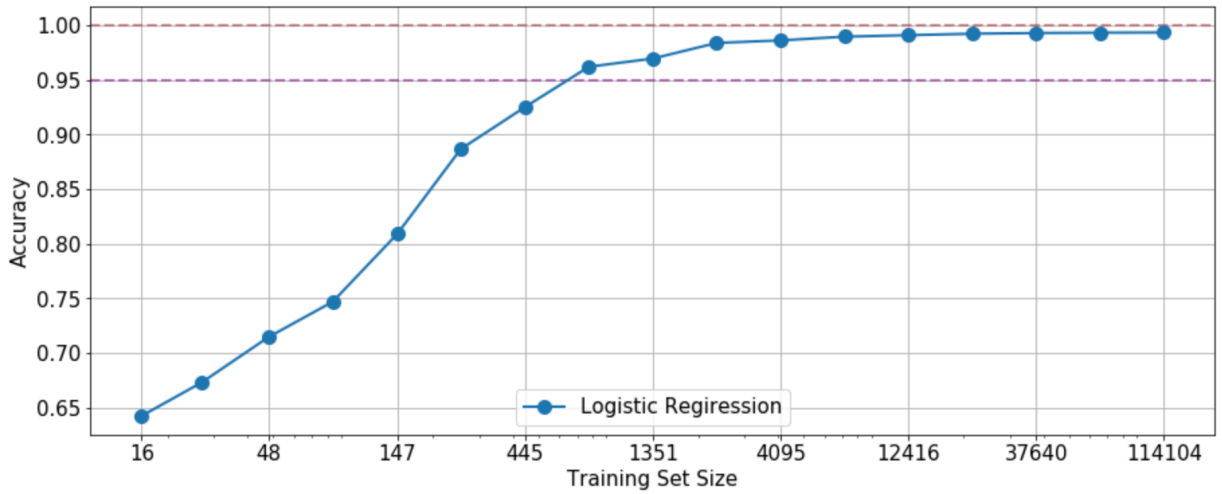

Рис. 1. Достигнутая точность логистической регрессией (N = 64).

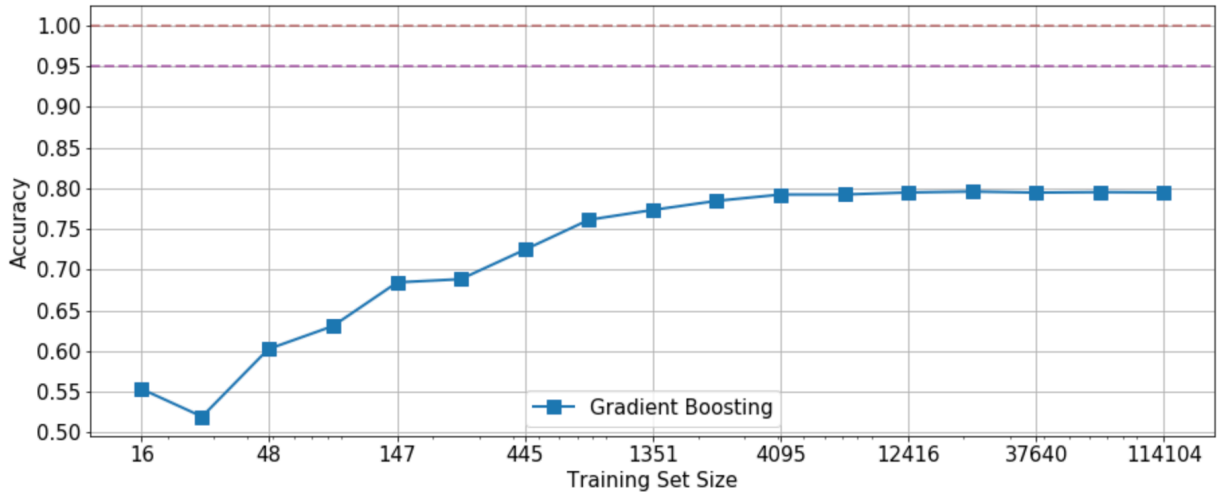

Рис. 2. Достигнутая точность градиентным бустингом (N = 64).

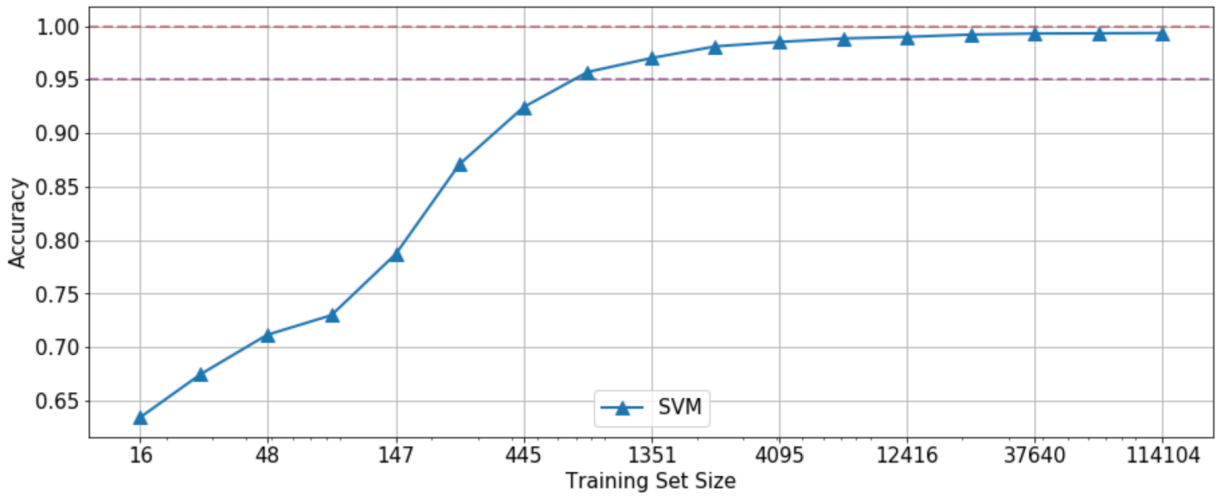

Рис. 3. Достигнутая точность методом опорных векторов (N = 64).

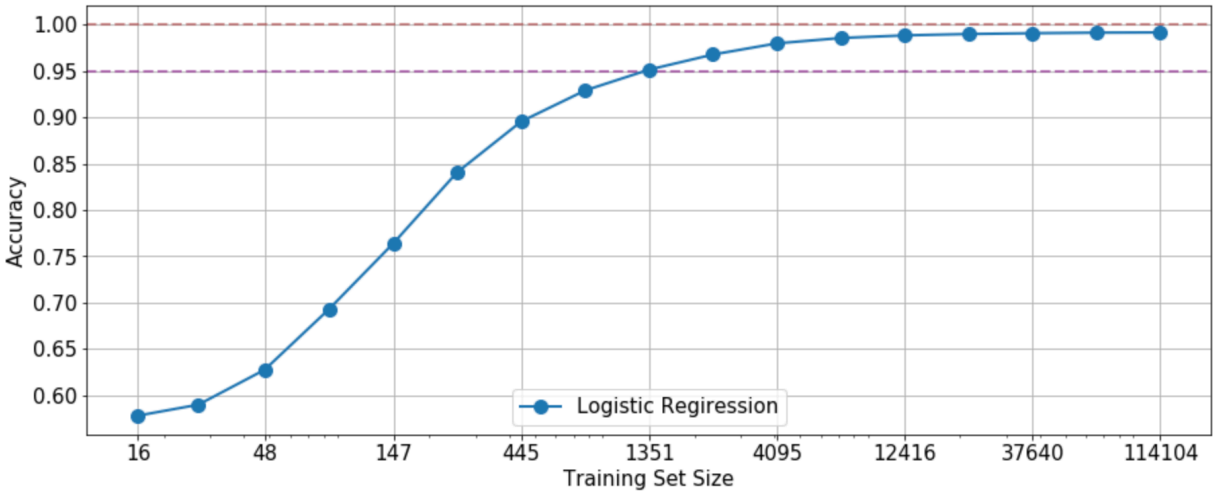

Рис. 4. Достигнутая точность логистической регрессией (N = 128).

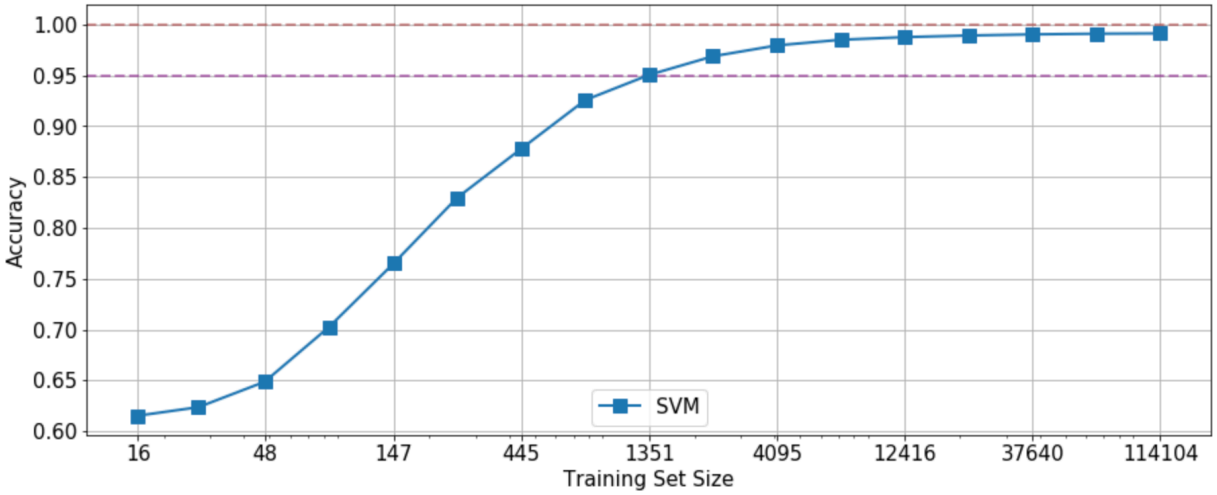

Рис. 5. Достигнутая точность методом опорных векторов (N = 128).

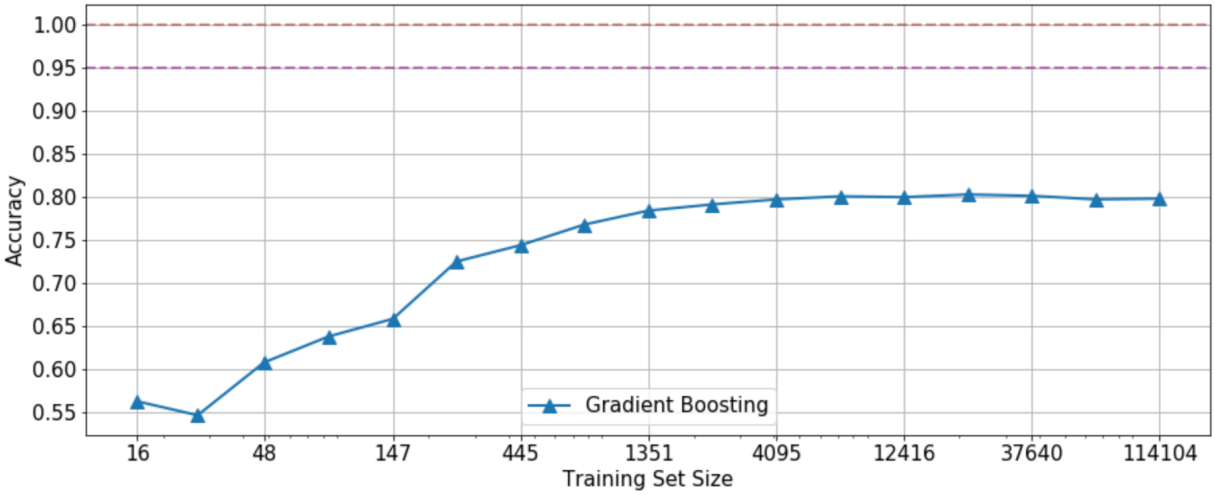

Рис. 6. Достигнутая точность градиентным бустингом (N = 128).

Зависимости показывают, что наименьшая длина выборки для достижения точности в 95% необходима для алгоритма логистической регрессии. Метод опорных векторов показал такие же результаты, что и логистическая регрессия,

однако требовал большего времени для обучения. Градиентный бустинг не достиг необходимой точности ни при каком значении длины выборки меньшей 10^5. В следующей таблице (таблица 1) представлены значения длин выборок, необходимые для достижения необходимой точности:

Таблица 1. Длины выборок необходимые для достижения точности 95%

Алгоритм	Длинна запроса 64 бит	Длинна запроса 128 бит
Логистическая регрессия	~ 700	~ 700
Метод опорных векторов	~ 1500	~ 1500
Градиентный бустинг	Точность не достигнута	Точность не достигнута

Зависимость точности предсказания от длинны запроса ФНФ приведена на рисунках 7 - 9. Обучение проводилось на выборке длинной 10^5 запросов и ответов. Точность подсчитывалась на выборке размером $9 \cdot 10^5$.

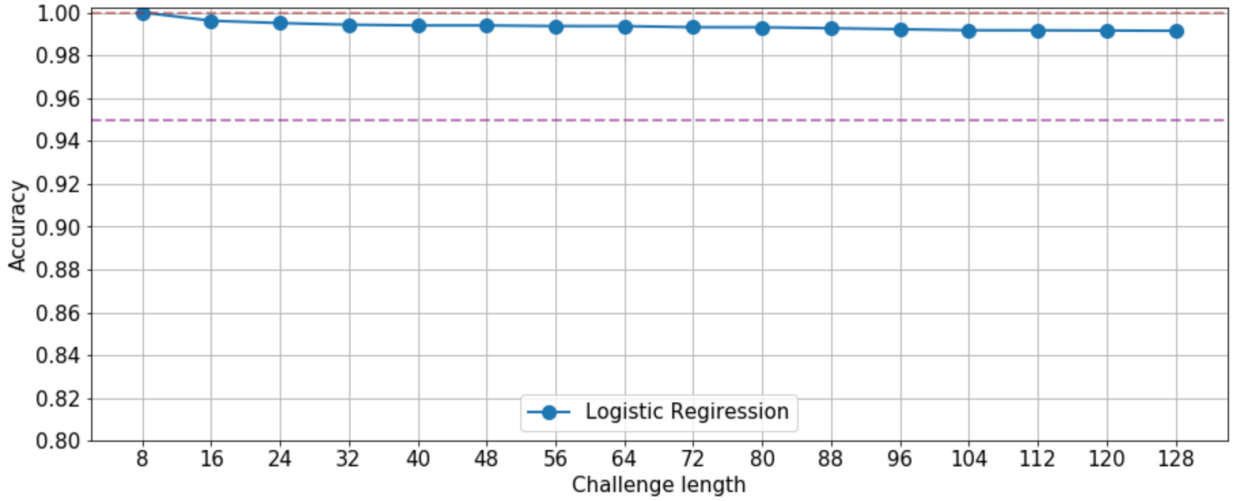

Рис. 7. Зависимость точности предсказания от длинны запроса для логистической регрессии.

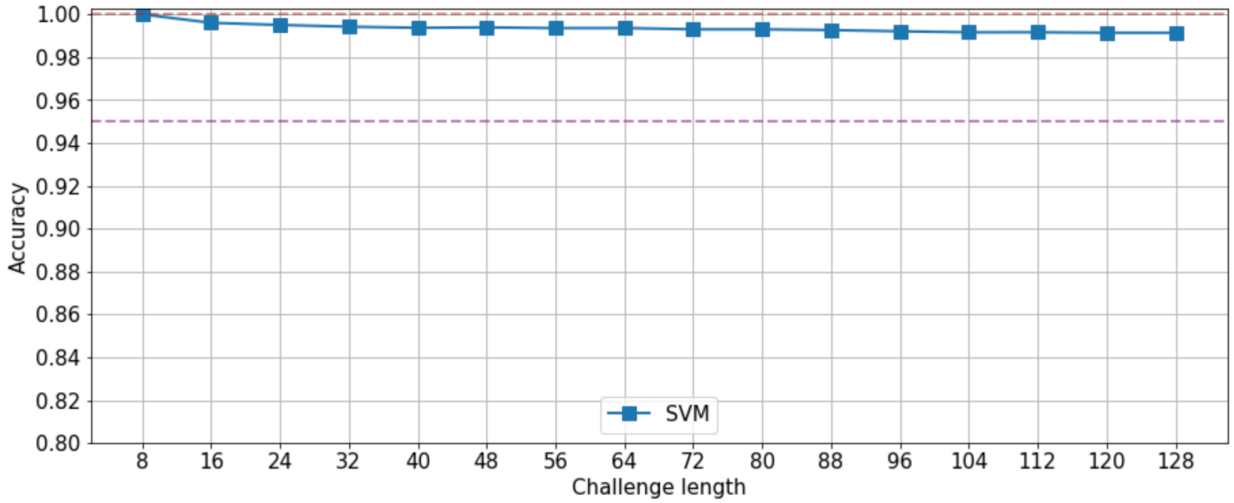

Рис. 8. Зависимость точности предсказания от длинны запроса для метода опорных векторов.

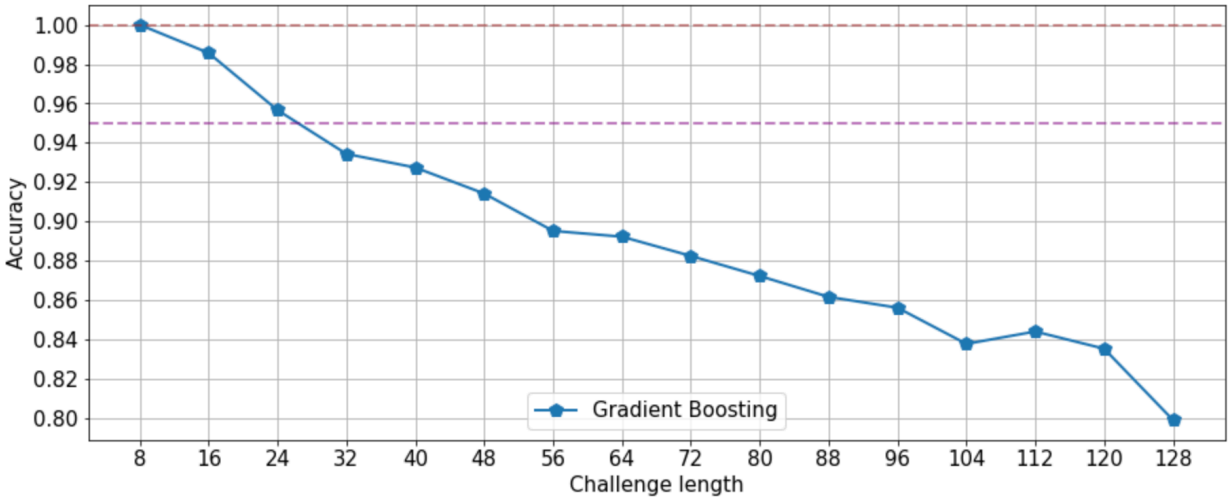

Рис. 9. Зависимость точности предсказания от длинны запроса для градиентного бустинга.

Зависимость показывает, что наилучшим алгоритмом для построения модели предсказания ответов ФНФ является логистическая регрессия. Точность такой модели почти не падает при увеличении длинны запроса. Также она эффективна в плане скорости обучения и предсказаний.

Метод опорных векторов показал точность не хуже, чем логистическая регрессия. Однако время обучения оказалось в разы больше, чем у модели выше.

Точность предсказания градиентного бустинга значительно падает с ростом длинны запроса. Это может быть связано с тем, что при большой длине запроса количество признаков большое и градиентный бустинг плохо справляется с такими задачами при относительно небольших объемах выборки.

4. Вывод

В результате работы были созданы модели для предсказания ответов ФНФ по запросам с помощью следующих алгоритмов: логистическая регрессия, метод опорных векторов, градиентный бустинг. Логистическая регрессия показана наилучшую эффективность как в точности предсказания, так и в скорости работы.