Московский авиационный институт

(национальный исследовательский университет)

Факультет информационных технологий и прикладной математики

Кафедра вычислительной математики и программирования

Лабораторная работа №1 по курсу «Искусственный интеллект»

(6-й семестр)

Студент: Сизонов А. А.

Группа: М8О-408Б

Дата:

Оценка:

Подпись:

**Москва, 2019**

Тема:

Azure ML Studio

**Задание:**

Ваша задача познакомиться с платформой Azure Machine Learning, реализовывав полный цикл разработки решения задачи машинного обучения, использовав три различных алгоритма, реализованные на этой платформе.

**Требования:**

1. Уникальность решения

2. Обоснованность выбора той или иной операции

3. В отчете должны быть указаны алгоритмы, которые применялись, результаты применения этих алгоритмов, а также скрины некоторых этапов обработки данных

Источники данных:

<https://vincentarelbundock.github.io/Rdatasets/csv/Ecdat/Computers.csv>

**Описание решения:**

База, выбранная мной хорошо подходит для задачи классификации. В качестве предсказываемой метки я выбрал “Premium” – является ли компьютер премиальным. Имеем два класса.

Алгоритмы:

1) Two-Class Averaged Perceptron

2) Two-Class Bayes Point Machine

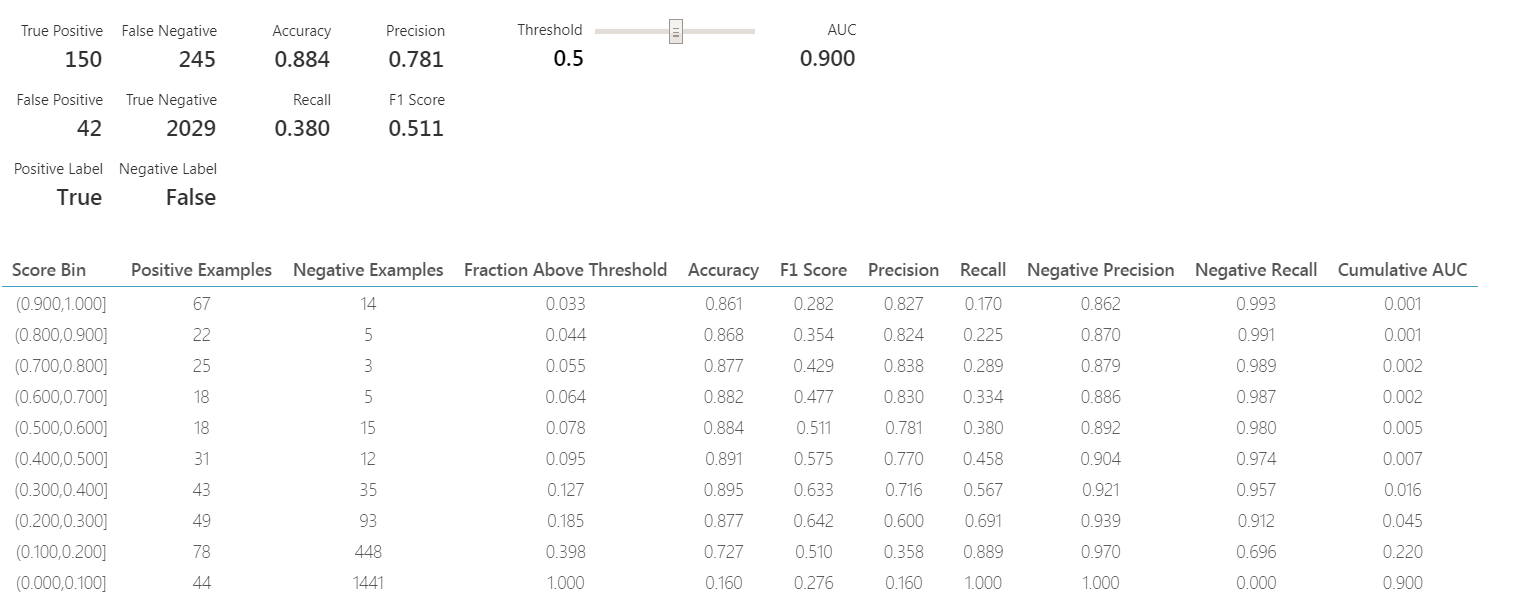
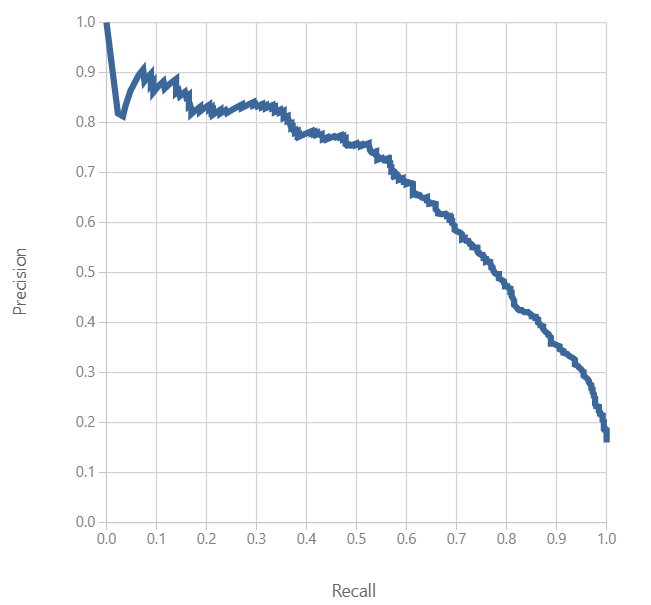
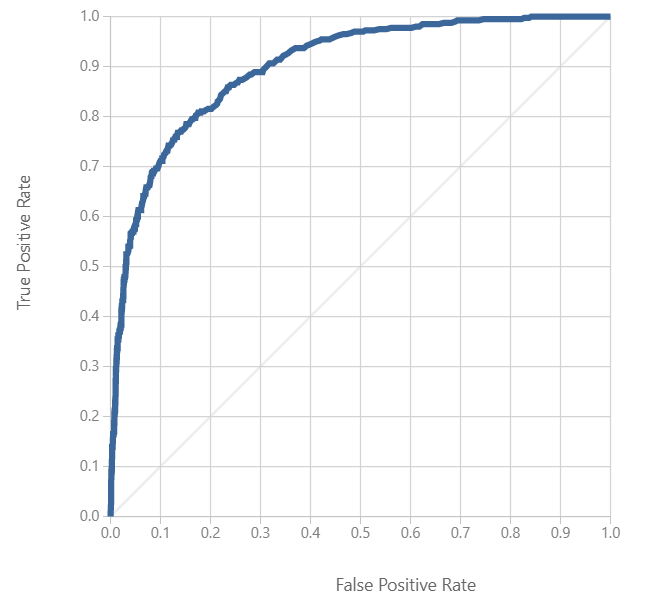
3) Two-Class Decision Forest

Сравнение происходило следующим образом:

1. Данных достаточно (разбиение 80/20).
2. Данных не хватает.
3. Переизбыток данных.

Переобучение и недообучение получалось только в вырожденных случаях.

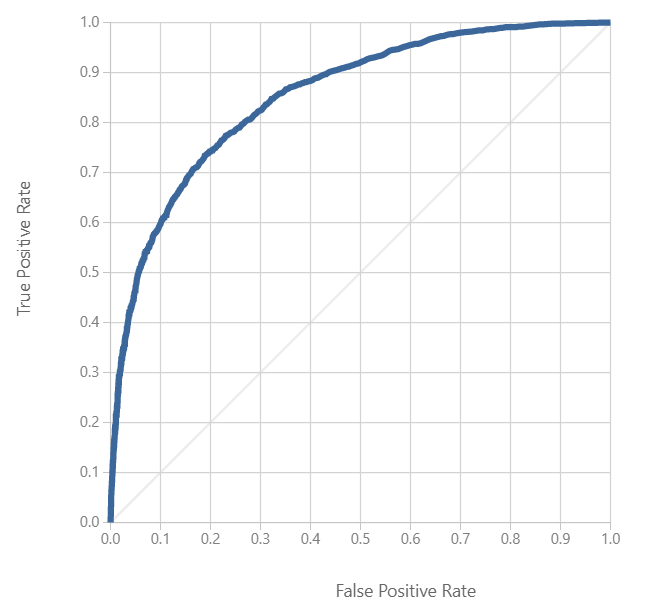
Работа персептрона при разбиении 80% к 20%:

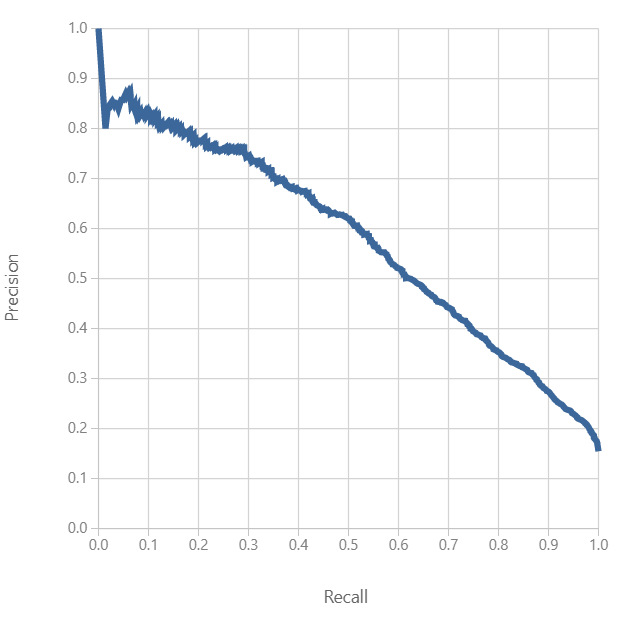


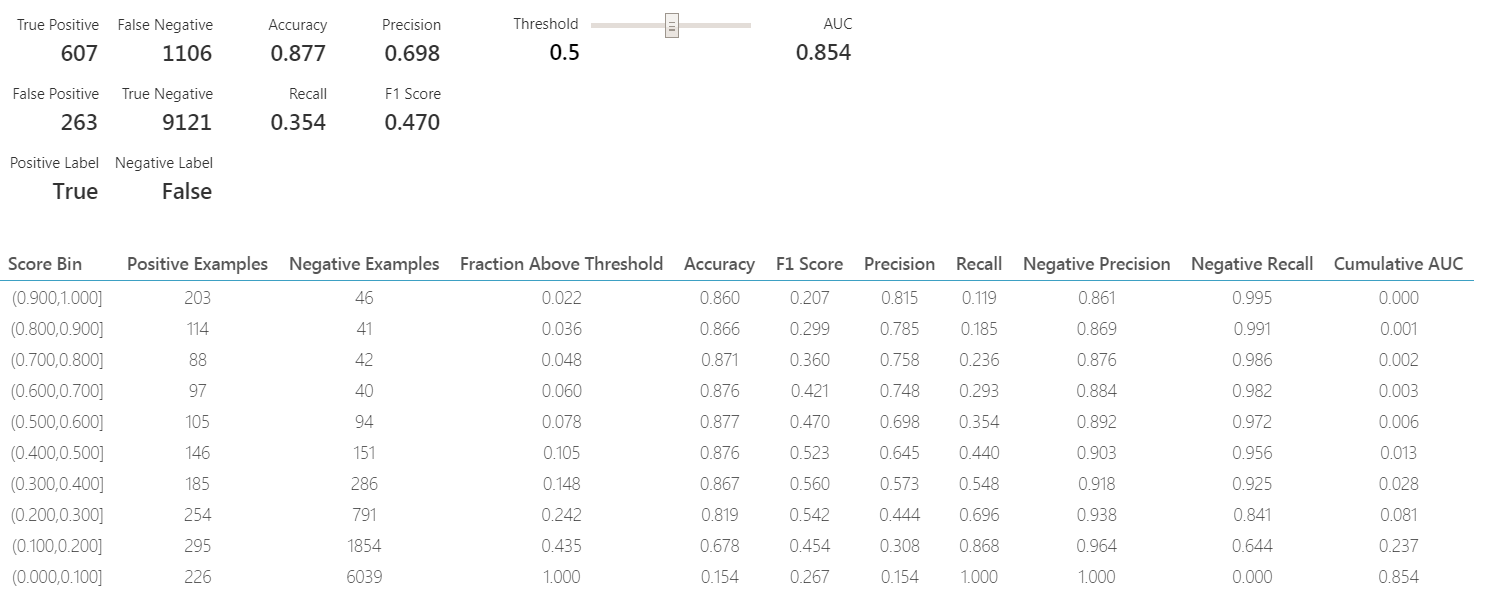
Как видно, средняя точность достигает 90%, что является хорошим показателем.

При переобучении и недообучении точность, что логично, падает, хоть и не сильно.

Работа персептрона при разбиении 1% к 99%:



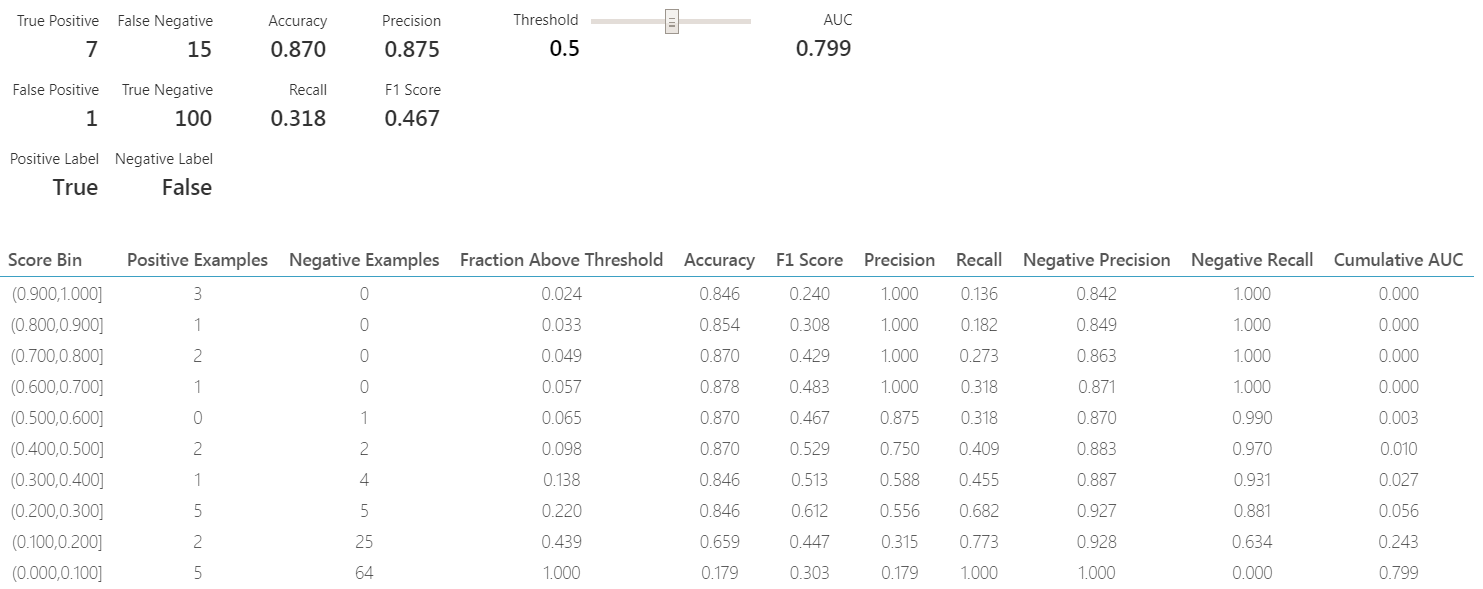
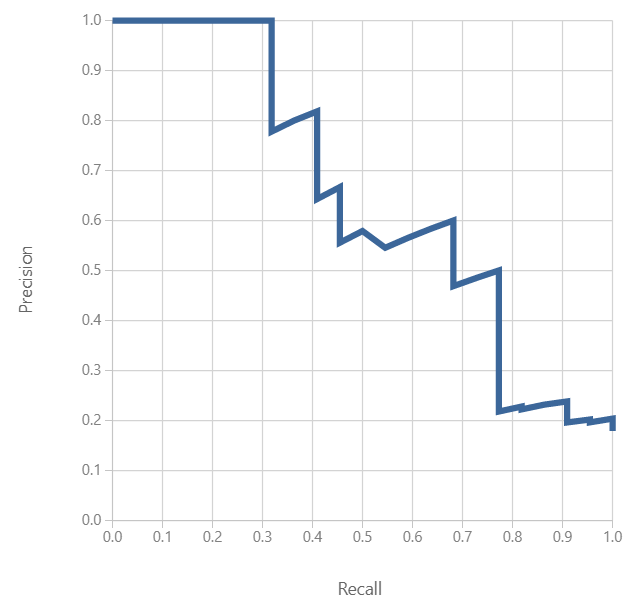
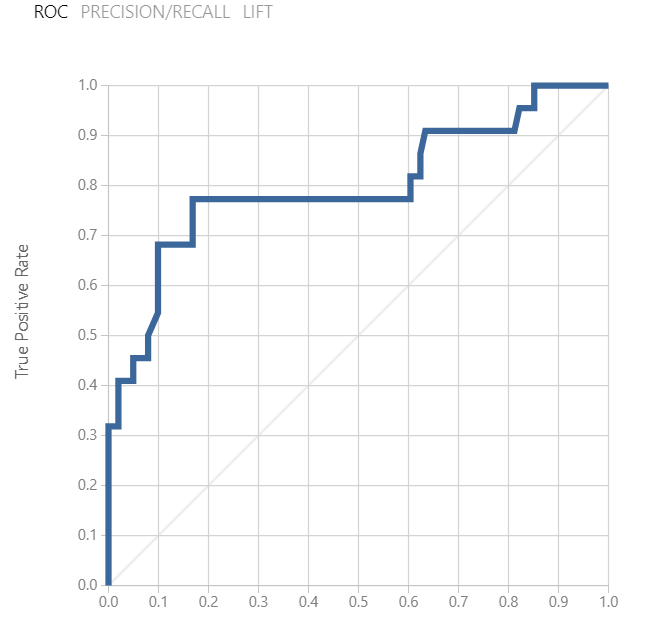




Средняя точность упала до 85.4%.

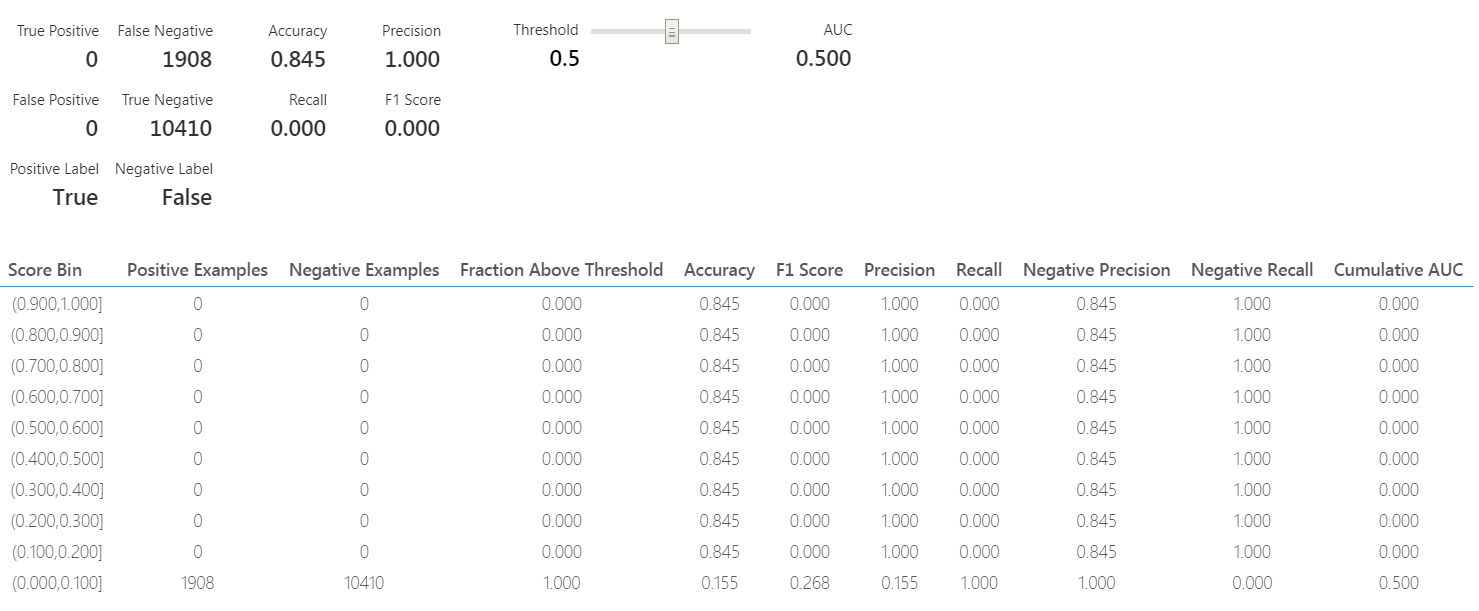
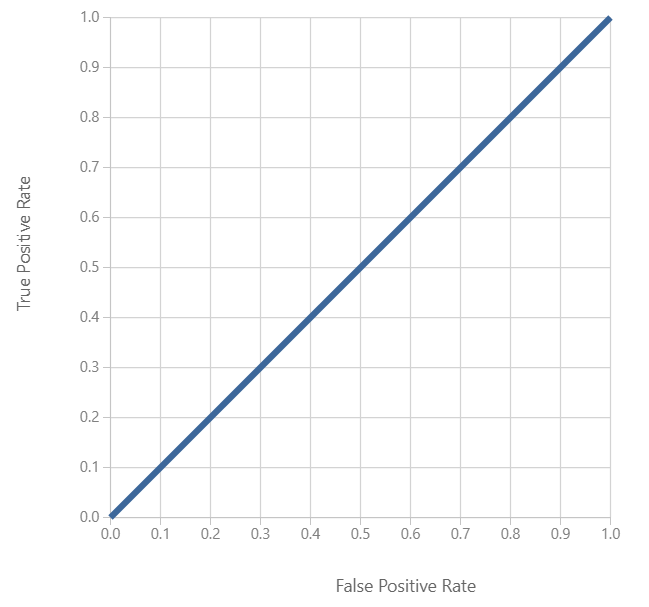
Малое падение точности связано с хорошей оптимизацией алгоритма и регулировкой.

Работа персептрона при разбиении 99% к 1%:



Точность упала до 79.9%. Это обусловлено переобучением.

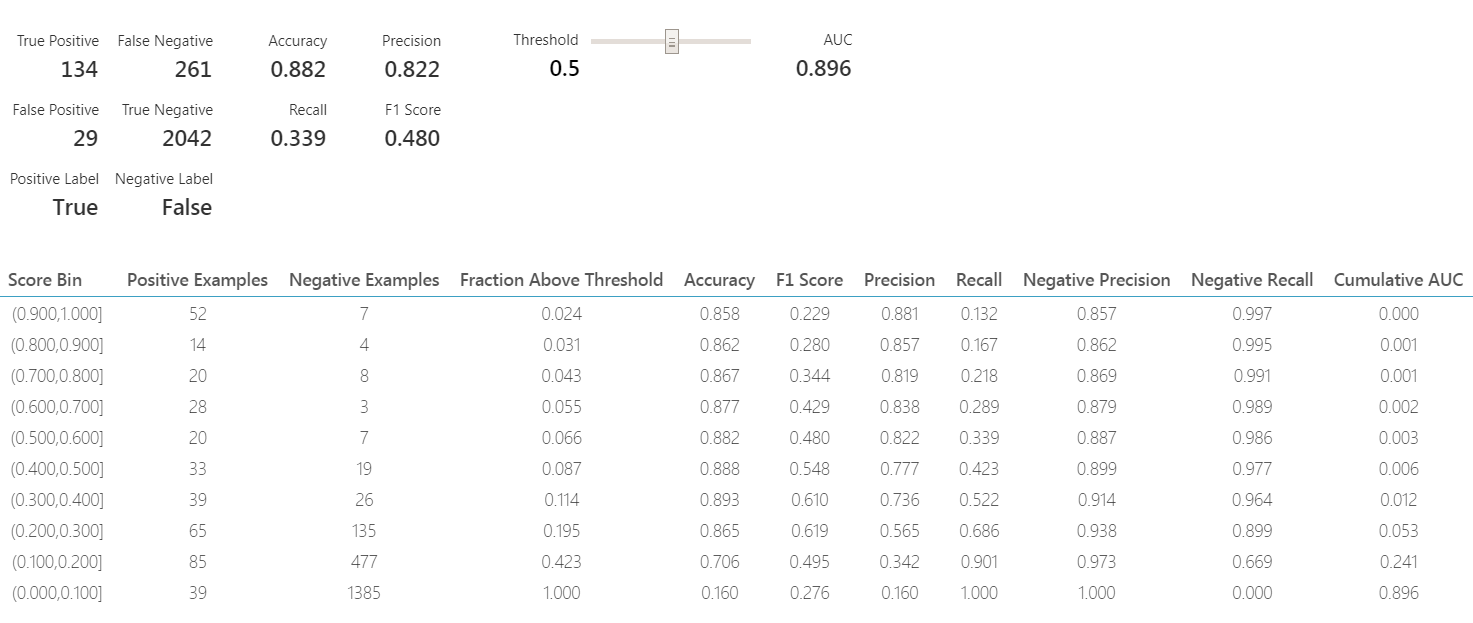
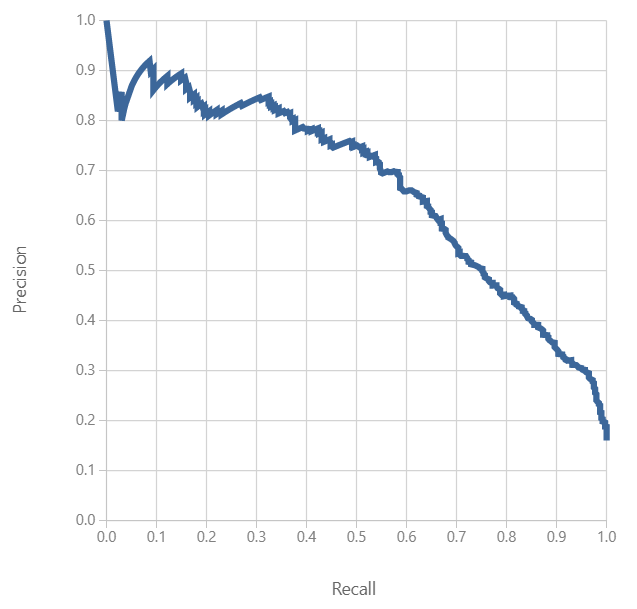
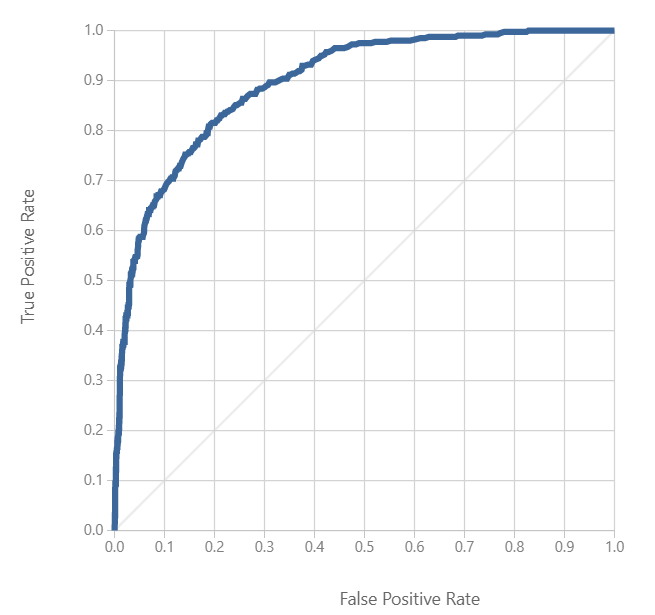
Разбиение 0.001/0.999: модель не может обучиться, так как ей не хватает данных.



При таком малом количестве данных для обучения сходимости не происходит и следовательно, вероятность угадать метку становится всего 50%.

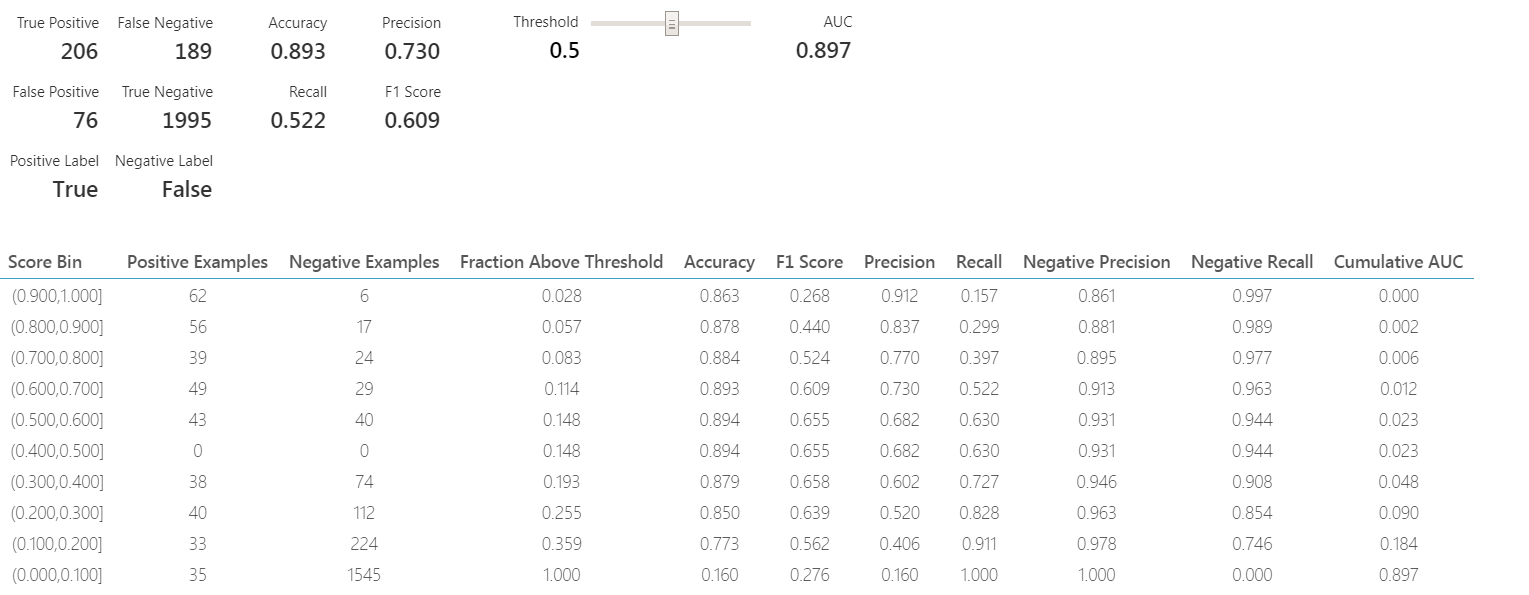
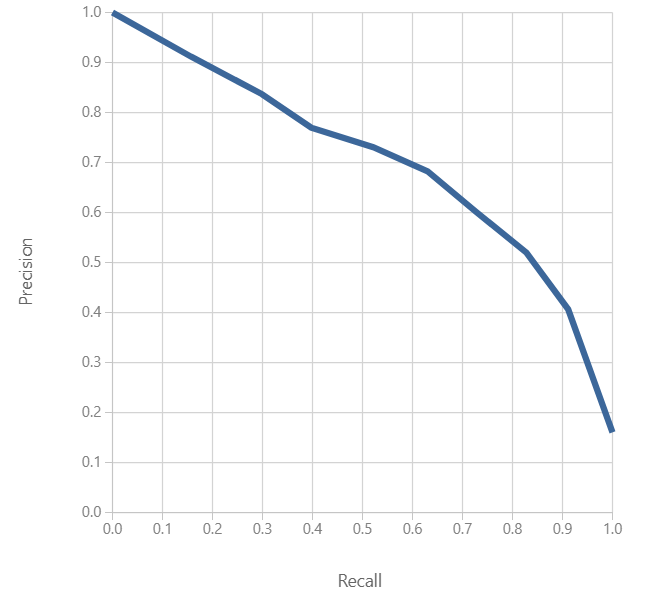
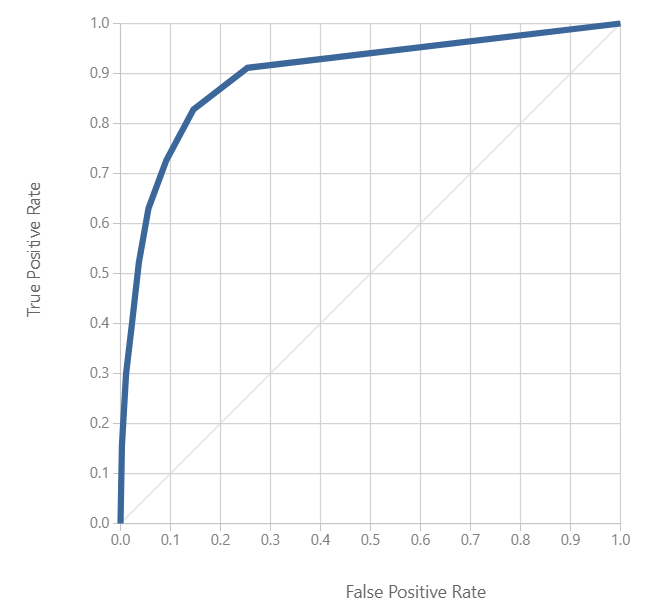
Для остальных алгоритмов обучения приведу результаты только для разбиение 80/20.

Байесовский метод (разбиение 80/20):



Средняя точность этой модели – 89.6%.

Алгоритм с использованием деревьев решений (разбиение 80/20)::



Средняя точность – 89.7%.

Из-за специфики данных самым точным алгоритмом оказался персептрон, но на другой базе, скорее всего, самым точным будет алгоритм с использованием деревьев решений, так как он более совершенный.

**Выводы:**

Благодаря этой лабораторной работе был получен навык работы с Azure Machine Learning Studio. Считаю его очень удобным, особенно для начинающих разбираться в машинном обучении. Сервис предоставляет множество алгоритмов обучения, интуитивно понятный интерфейс, возможность загружать свою собственную базу, множество вариантов визуализации того, что происходит в процессе обучения модели, модульность и многое другое.