ISA 419: Data Driven Security

17 - Touch Analytics Lab

Fadel M. Megahed

Endres Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Automated Scheduler for Office Hours

Fall 2022

Outline

1 Preface

2 An Overview of the Touchalytics Paper (Frank et al. 2012)

Learning Objectives for Today's Lab

Objectives

- Examine an Android Phone's behavioral touch feature dataset
- Replicate the analysis performed by a InfoSec research paper
- Report the code and results in an understandable manner to a general audience

Outline

1. Preface

2. An Overview of the Touchalytics Paper (Frank et al. 2012)

Preface

The paper by Frank et al. (2012) is one of the seminal papers, where smart phone touch features are engineered and used for continuous authentication. The paper has over 600 citations, and provides a website that contains:

- The raw data (do not forget to examine the readme file)
- The extracted features (see the readme)
- Matlab Script File for how the features were engineered from the raw data file
- PDFs of their published work; for our lab, we are interested in the Touchalytics Paper

Paper's Motivation [1]

- There is limited work on continuous authentication for touchscreen devices
- Use of smart phone devices typically involves "atomic navigation behavior", which consists of simple and short movements
- Is it possible to authenticate users while they perform basic navigation steps on a touchscreen device and without any dedicated and explicit security action that requires attention from the user?

Paper’s Motivation [2]

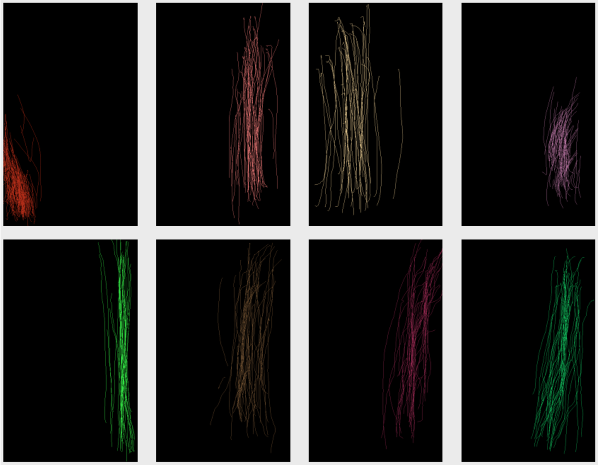


Figure 1: Strokes from eight different users, each reading three different texts on an Android Phone.

Paper's General Idea and Goals [1]

There are two phases for learning and classifying touch behavior.

- **Enrollment Phase:** Phase where the system relies on a conventional authentication approach (e.g., password). The authors distinguish between two different "trigger-actions":
  - Sliding horizontally over the screen; for example, to browse through images
  - Sliding vertically over the screen to move conten up and down
  - Note that the authors decided to use only on on **single touch gestures**
- **Continuous Authentication Phase:** During this phase, the system continuously tracks all strokes and the classifier estimates if they were made by the legitimate user. For t consecutive negative classification results, the system resorts back to the initial entry-point based authentication method and challenges the user.

Paper's General Idea and Goals [2]

Study Goals

The main goal of their study is to analyze how robustly our proposed framework can distinguish users from each other.

- What is the probability of rejecting a legitimate user?
- What is the probability of accepting an attacker?
- How long does the classifier need to make an authentication decision?
- How robust is the classification within one session, across multiple sessions, and after one week?

Experimental Protocol

- Sessions 1-3 occured during the same day and seperated by a questionnaire. This is followed by a second phase, where users were asked to spot differences between images.
- Sessions 6-7 occured one week post the sessions in the previous bullet.

Touch Analytics [1]

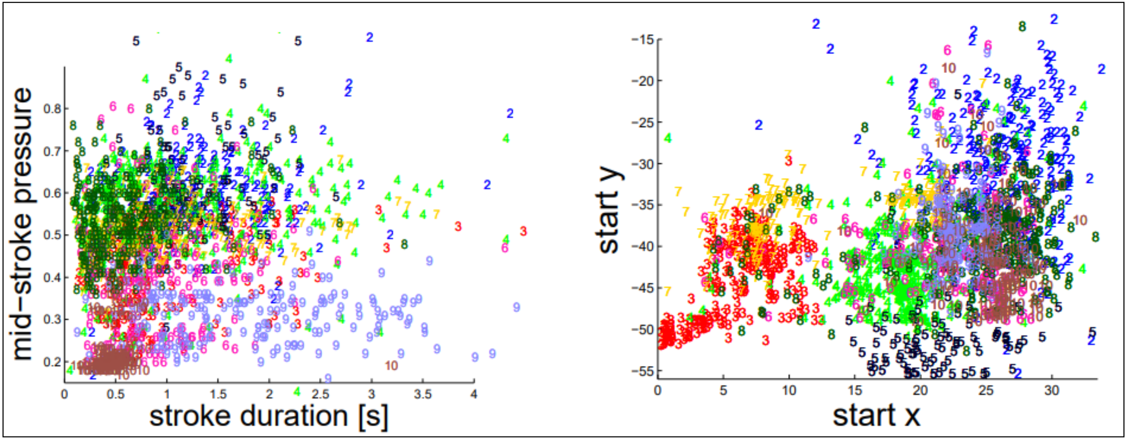


Figure 2: Stroke features projected on a 2D-subspace. The user ID is given as a colored number.

Touch Analytics [2]

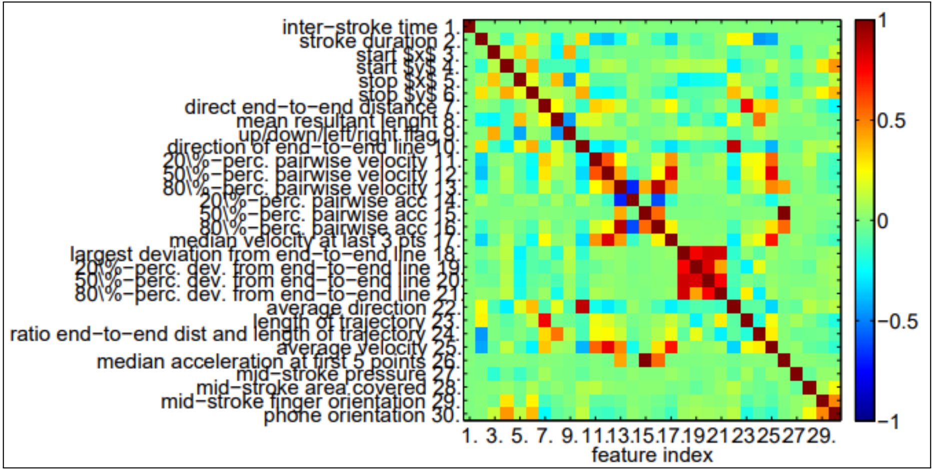


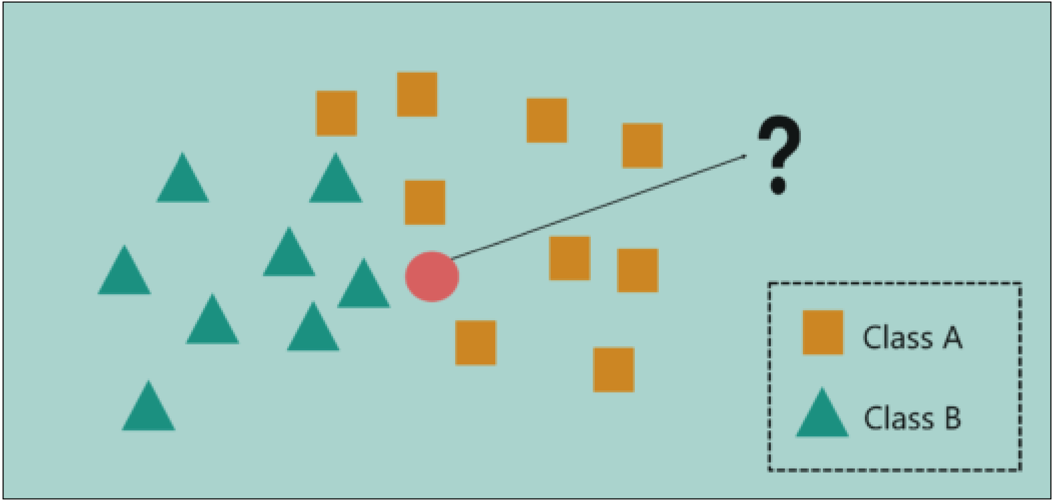
Figure 3: Correlation among the 30 features.

Classification Schemes: kNN [1]

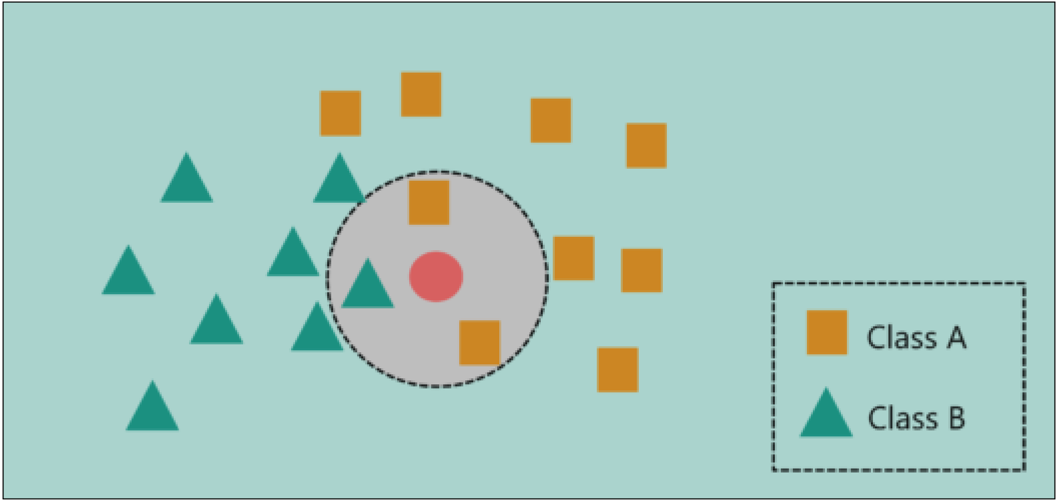
The kNN (k-Nearest Neighbors) algorithm has the following features:
- Supervised learning algorithm
- Utilizes feature similarity
- Lazy algorithm since it uses the entire training data to make a decision instead of coming up with a discrminative function

The following kNN Slides are based on the edureka.co Blog.

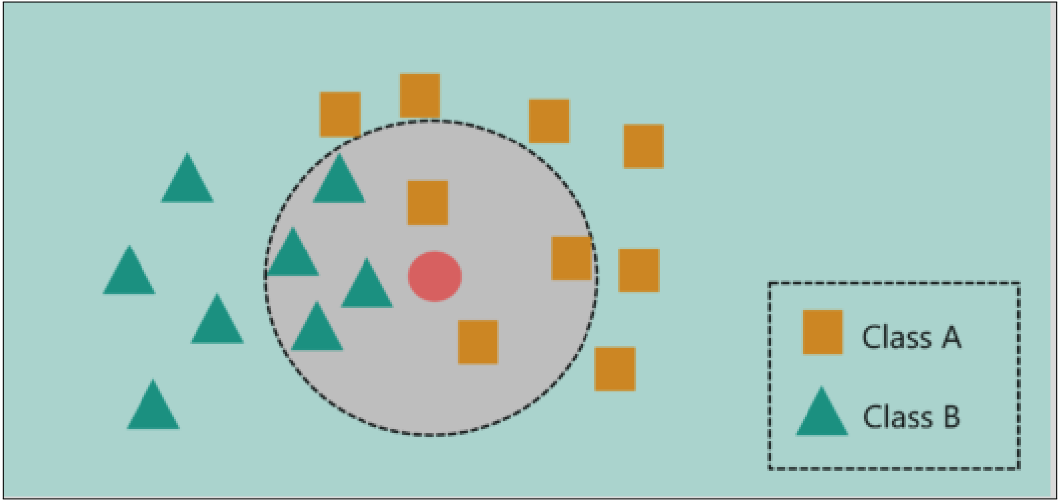
Classification Schemes: kNN [2]



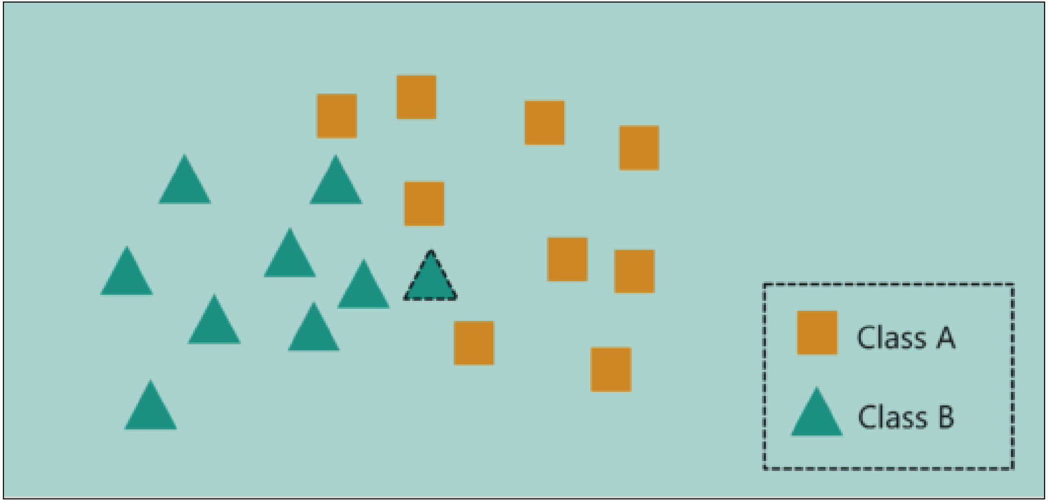
Classification Schemes: kNN [3]



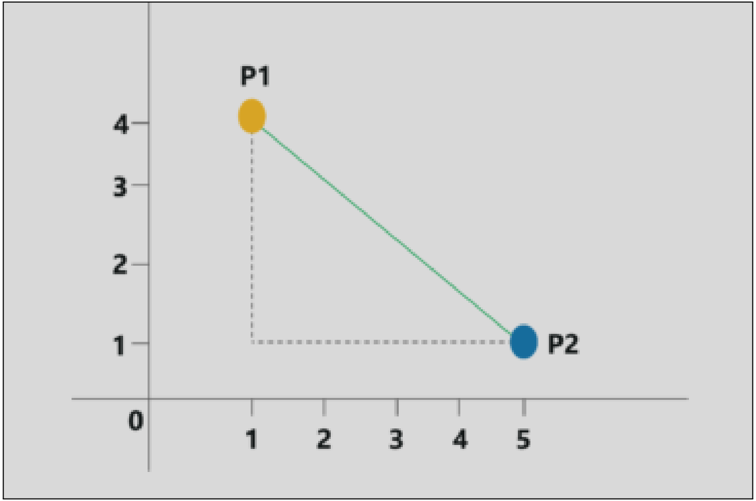
Classification Schemes: kNN [4]



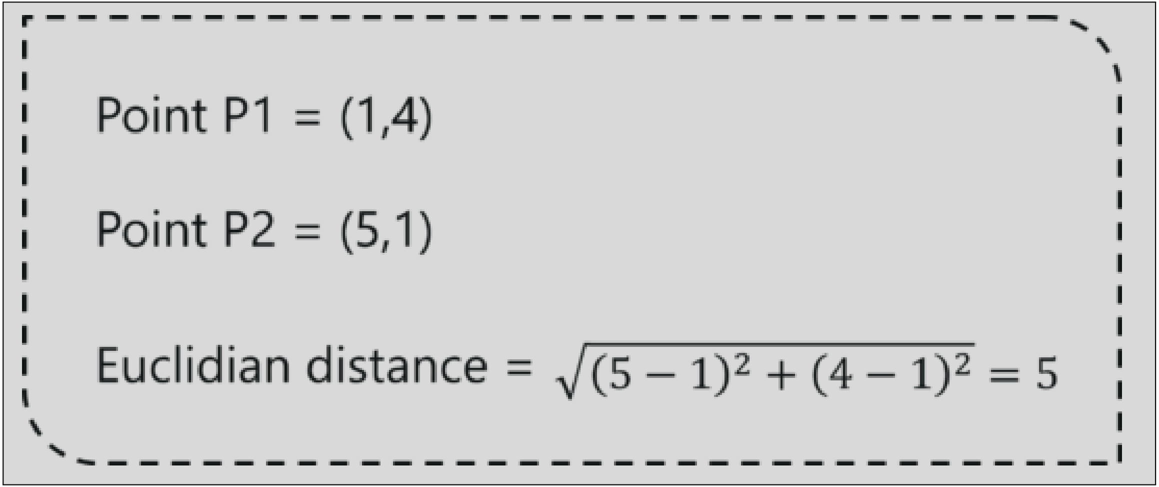
Classification Schemes: kNN [5]



Classification Schemes: kNN [6]



Classification Schemes: kNN [7]



Point P1 = (1,4)

Point P2 = (5,1)

Euclidian distance = $\sqrt{(5-1)^2 + (4-1)^2} = 5$

Classification Schemes: kNN [8]

Pseudo Code

- **Calculate $D(x, x_i)$, where $i = 1, 2, ..., n$ and D is the Euclidean measure between the data points.**
- **The calculated Euclidean distances must be arranged in ascending order.**
- **Initialize k and take the first k distances from the sorted list.**
- **Figure out the k points for the respective k distances.**
- **Calculate k_i, which indicates the number of data points belonging to the ith class among k points i.e. $k \geq 0$**
- **If $k_i > k_j \forall i \neq j$; put x in class i.**

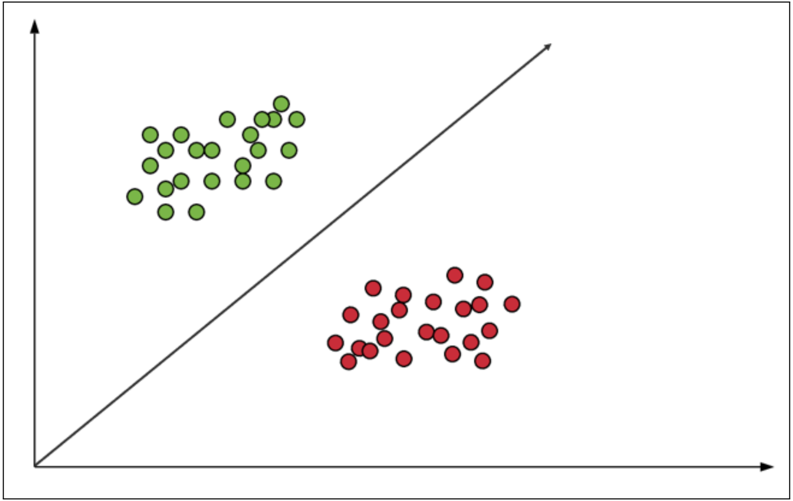
Note that in the paper of Frank et al. (2012), they have used $k \in [1, 3, 5, 7]$ to avoid ties.

Classification Schemes: SVM [1]

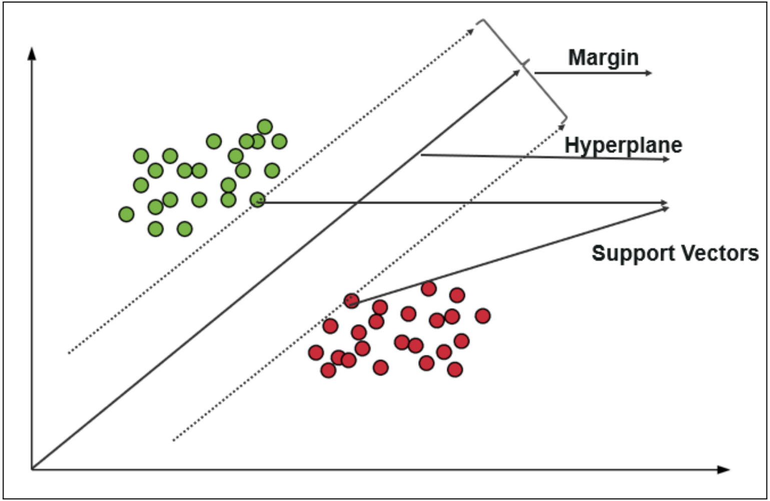
In this paper, they used it for binary classification. The used a RBF for the kernel, which means that they assumed that the data is not linearly seperable.

The following SVM Slides are based on the edureka.co Blog.

Classification Schemes: SVM [2]

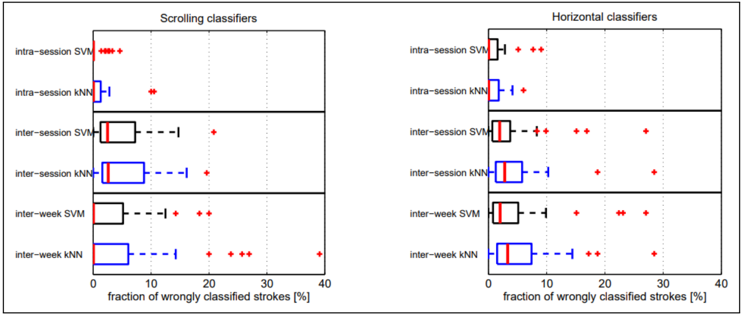


Classification Schemes: SVM [3]



Lab Deliverables [1]

For the purposes of Lab 04, I would like you to replicate the analysis performed by Frank et al. (2012). You can start with the The extracted features and simplify the problem into a binary classification problem, where you will need to recode the data at least 40 times such that for a given user we compare him/her to everyone else.



Lab Deliverables [2]

Rubric

- A student will score 80% on the assignment if they can successfully use both kNN and SVM to distnguish one user of their choice from the rest of the group based on the inter-week sessions.
- A student will score 100% on the assignment if they can successfully use both kNN and SVM to distnguish one user of their choice from the rest of the group based on three different training approaches: (a) intra-session, (b) inter-session, and (c) inter-week sessions.
- A student will score 120% on the assignment if they can successfully use both kNN and SVM to distnguish all users from the rest of the group based on three different training approaches: (a) intra-session, (b) inter-session, and (c) inter-week sessions.

Lab Deliverables [3]

This lab assignment is due on Oct 31, 2022. Class time on April 1st will be used to answer some initial questions pertaining to this assignment. You can work individually or in a group of two. Your submission should include: (a) your Notebook, and (b) a CSV file with results from your code.

References [1]

Frank, Mario, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2012. "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication." *IEEE Transactions on Information Forensics and Security* 8 (1): 136–48.

ISA 419: Data Driven Security

17 - Touch Analytics Lab

Fadel M. Megahed

Endres Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Automated Scheduler for Office Hours

Fall 2022