

LISTEN CAREFULLY

Acoustic Keylogging Using Machine Learning



Scott Stevenson

@scottastevenson

scottstevenson88@gmail.com

Navid Shekoufa

@n6599

shekoufa.navid@gmail.com

I. KEYLOGGING

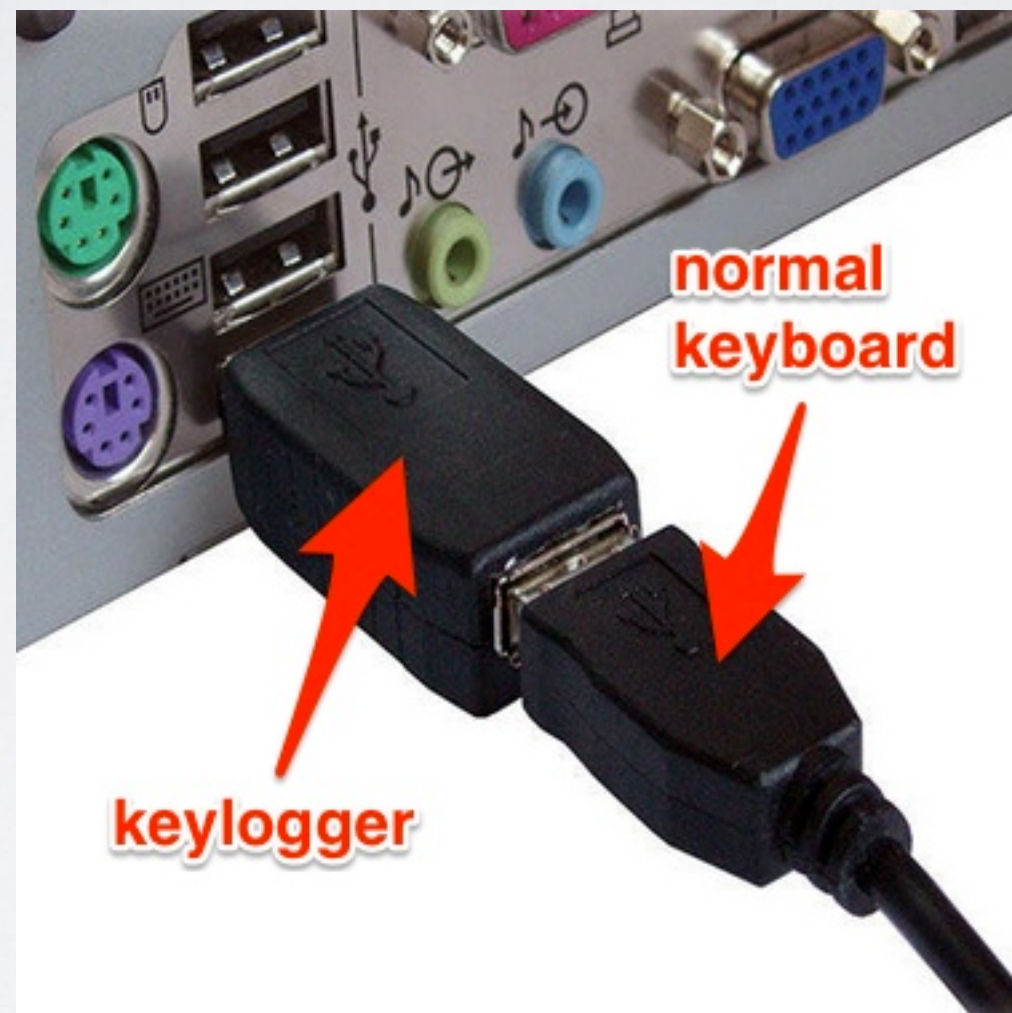
A **keylogger** is a type of surveillance software/hardware that records keystrokes a user makes.

They are often employed with the intention of stealing sensitive information such as **passwords** and **credit card numbers**.

SOFTWARE KEYLOGGERS

- Incredibly common in the 2000s - 70% of enterprises infected with keyloggers in 2008
- Completely undetectable in some OS's
- Virus protection got better - OS's more resilient
- But are still definitely out there...

HARDWARE KEYLOGGERS



HARDWARE KEYLOGGERS







Three researchers (**Li Zhuang, Feng Zhou, J. D. Tygar**) demonstrated in their paper “Keyboard Emanations Revisited” that they could recover **96% of typed characters** from a **10 minute recording** of keyboard sounds.

An improvement on the seminal work by **Asonov** and **Agrawal**.

2. MACHINE LEARNING

MACHINE LEARNING

- What is Machine Learning?
- Major types of Machine Learning:
 - Reinforcement Learning
 - Unsupervised learning
 - Supervised Learning

REINFORCEMENT LEARNING

- Inspired by behaviourist psychology
- Applications



I learned to ride with RL...

DATASETS

Play golf dataset

Independent variables				Dep. var
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
sunny	85	85	FALSE	Don't Play
sunny	80	90	TRUE	Don't Play
overcast	83	78	FALSE	Play
rain	70	96	FALSE	Play
rain	68	80	FALSE	Play
rain	65	70	TRUE	Don't Play
overcast	64	65	TRUE	Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	75	80	FALSE	Play
sunny	75	70	TRUE	Play
overcast	72	90	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play

UNSUPERVISED LEARNING

- Describes hidden structures
- Applications



SUPERVISED LEARNING

- Infers a function from labeled training data
- Training is involved
- Applications



ACOUSTIC KEYLOGGER

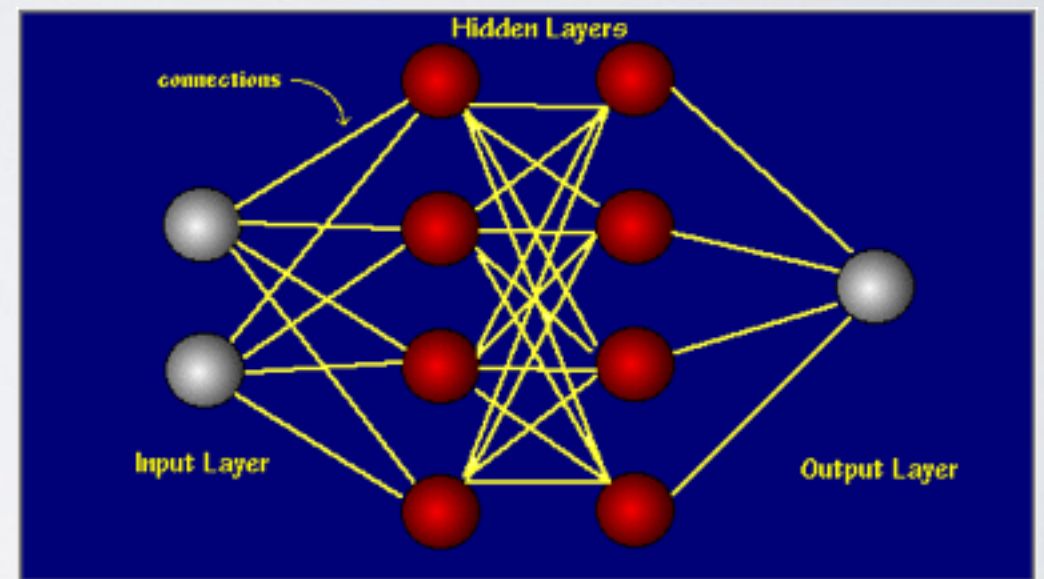
- Speech recognition is mostly a supervised learning process
- What would be the dataset in our demonstration?
- What would be the learning process?

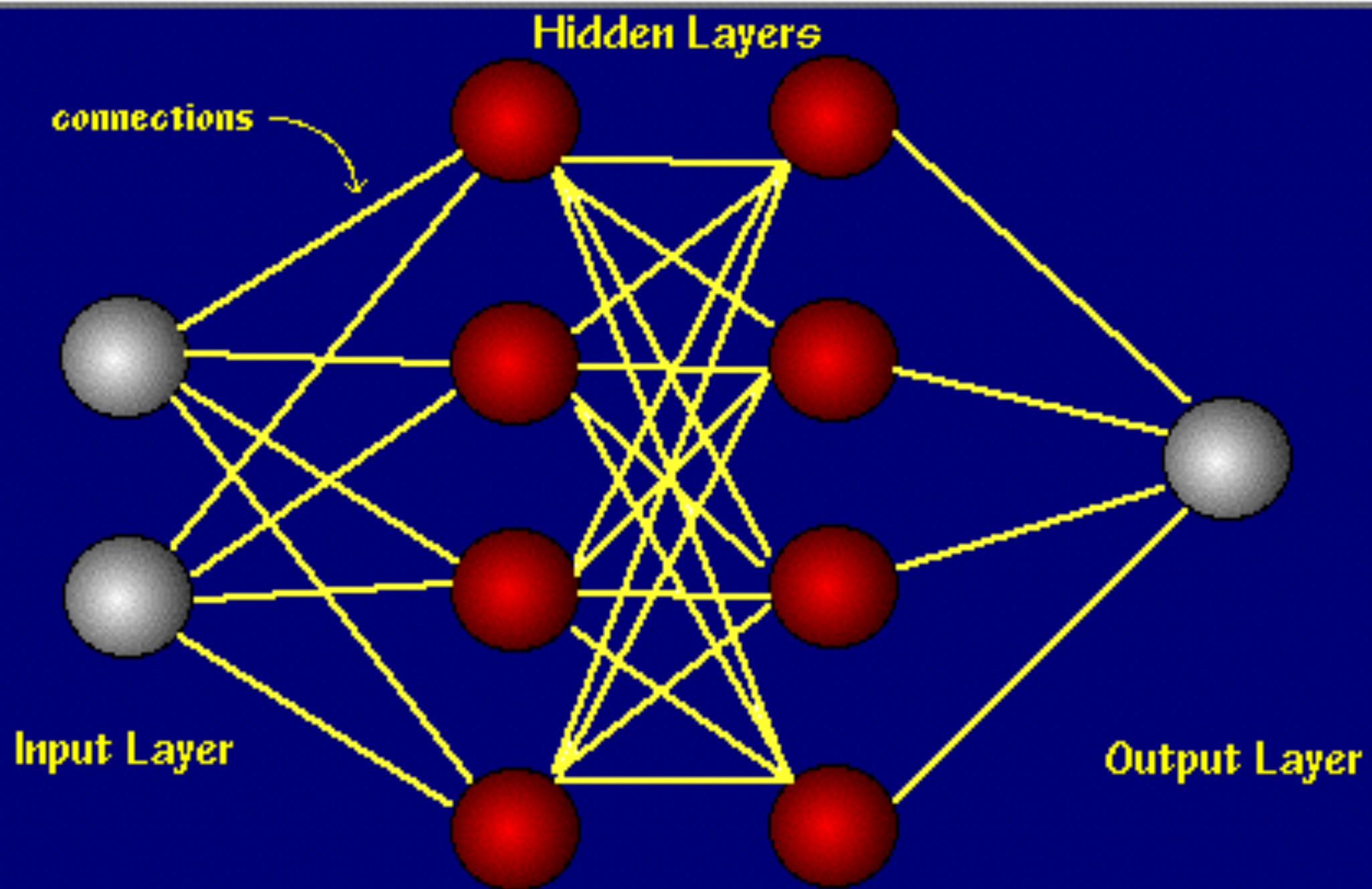
NEURAL NETWORKS

- Loosely modelled after the neuronal structure of the mammalian cerebral cortex
- A large ANN might have hundreds or thousands of processor units
- A mammalian brain has billions of neurons
- Complex mathematics involved
- Can rather easily gain an operational understanding of the operation

NEURAL NETWORKS

- Typically organized in layers
- Each node in each layer has an **activation function**
- Patterns are presented to the network via the **input layer**
- Actual processing is done in the **hidden layer**
- processing is done via a system of weighted **connections**
- ANNs contain some form of **learning rule** to update the weights
- ANNs **learn by example**, as a child learns to recognize dogs from **examples** of dogs

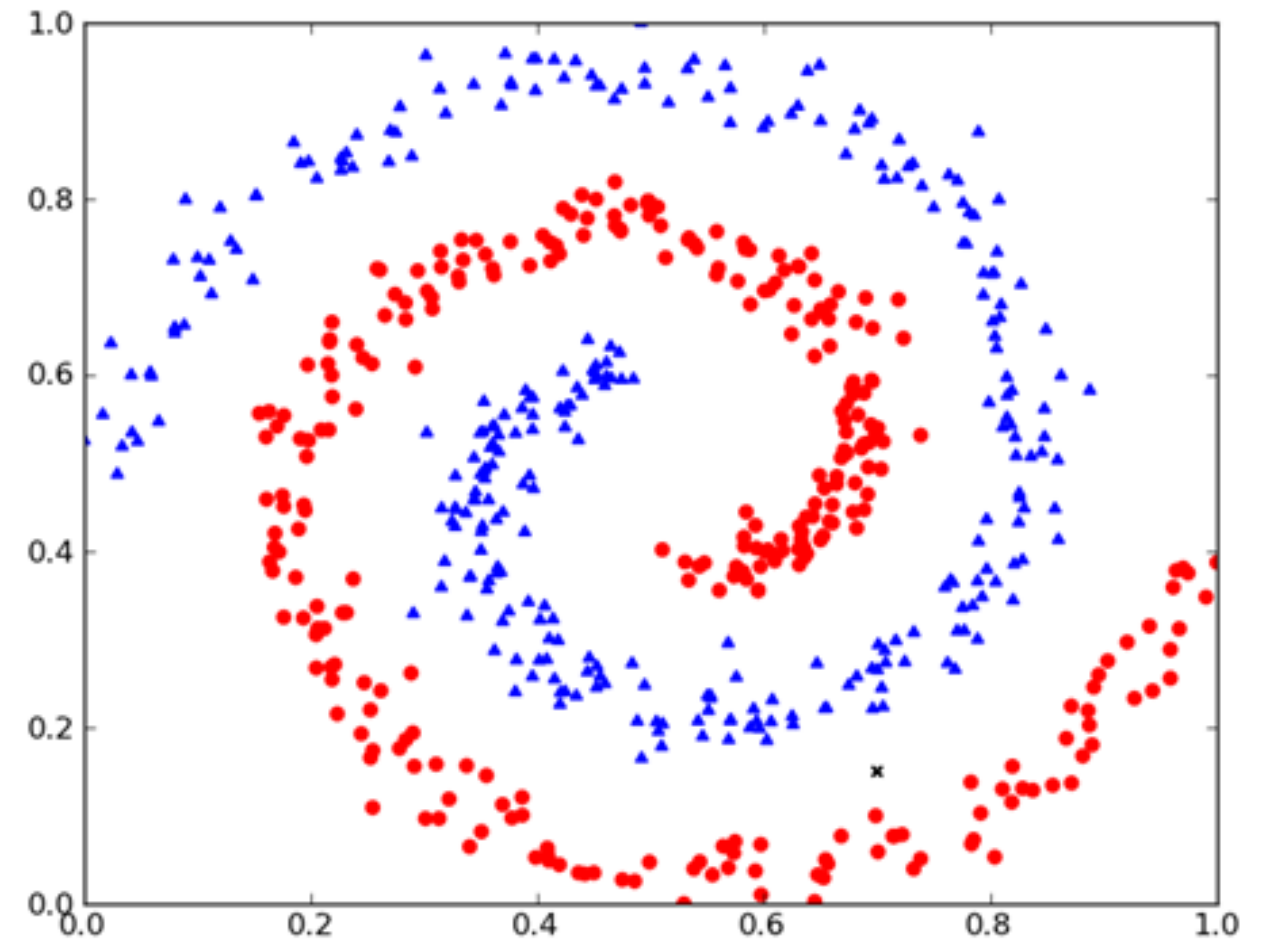
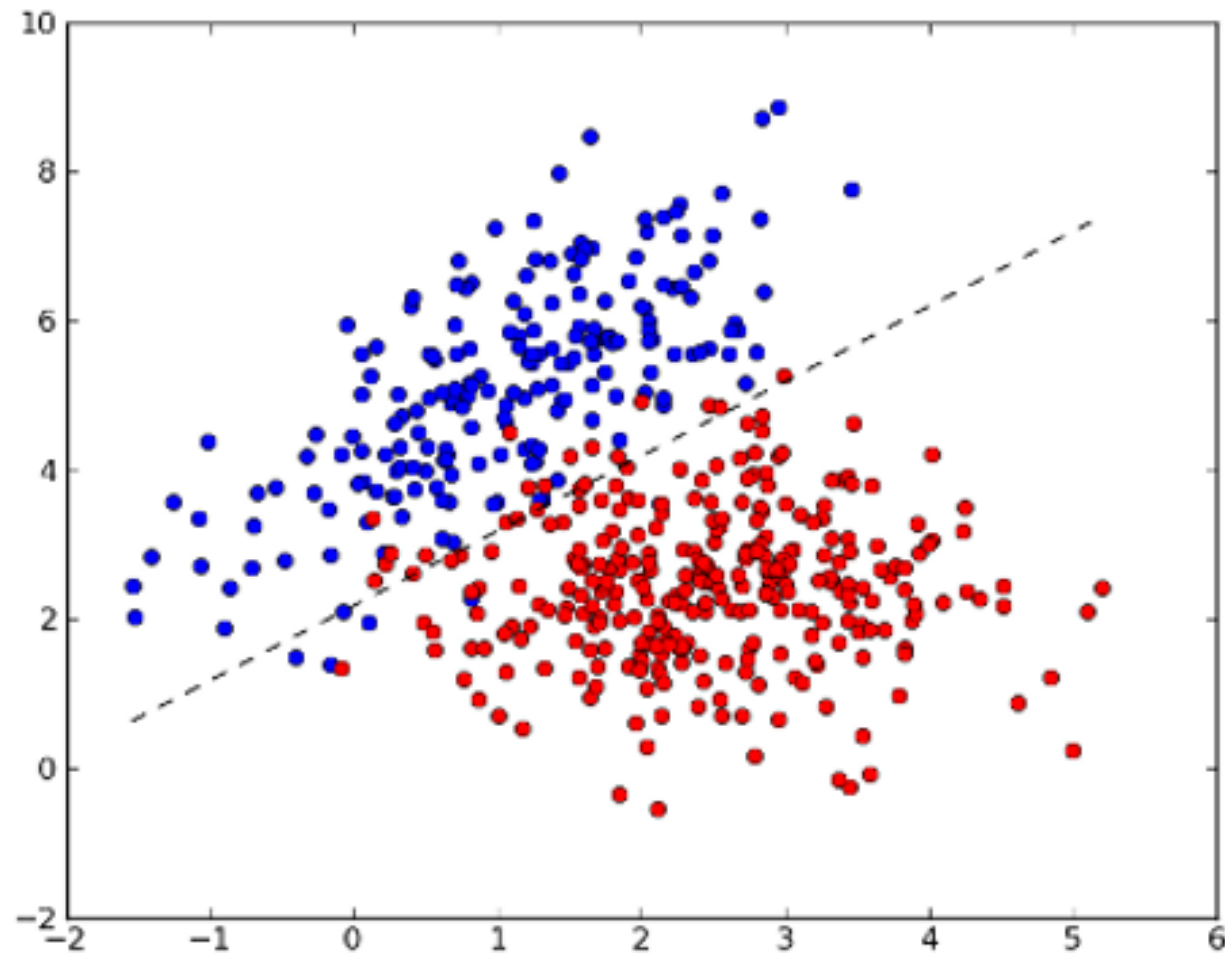




CHALLENGES

- What are the challenges?
 - Assumptions = Deviation from real world
 - Bias
 - Keeping the model simple

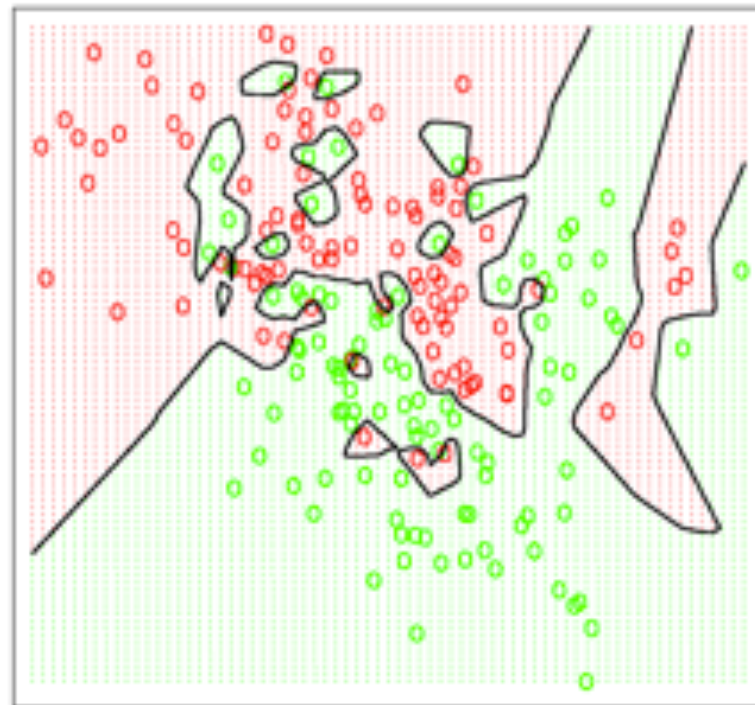
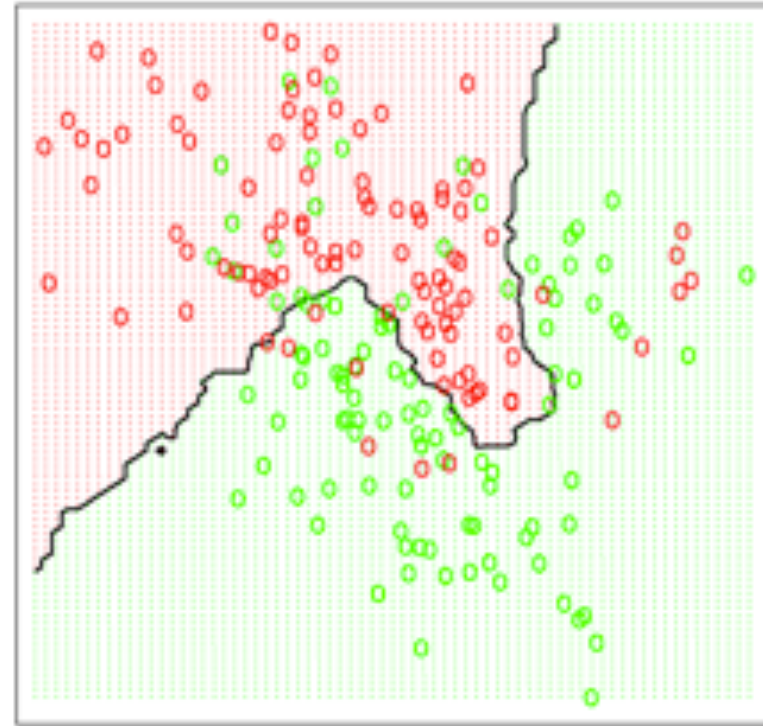
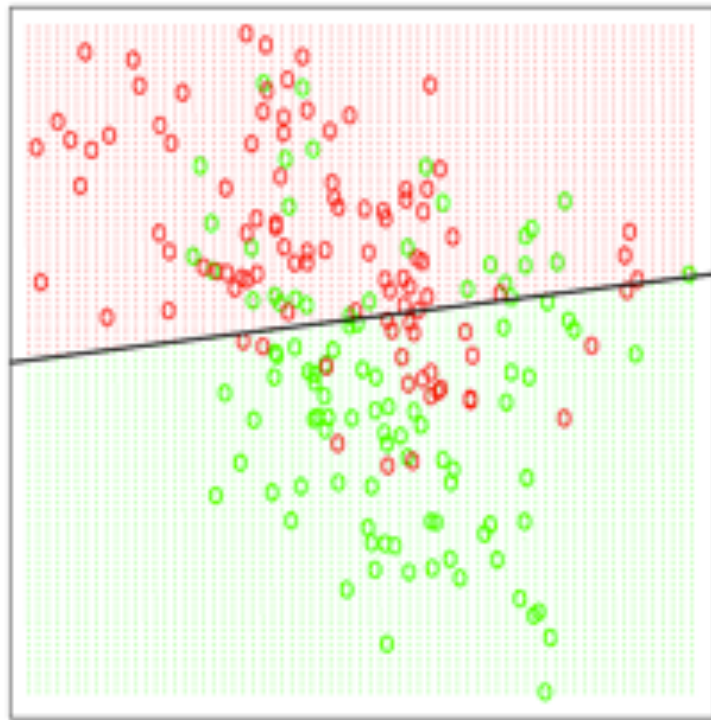
CHALLENGES



CHALLENGES

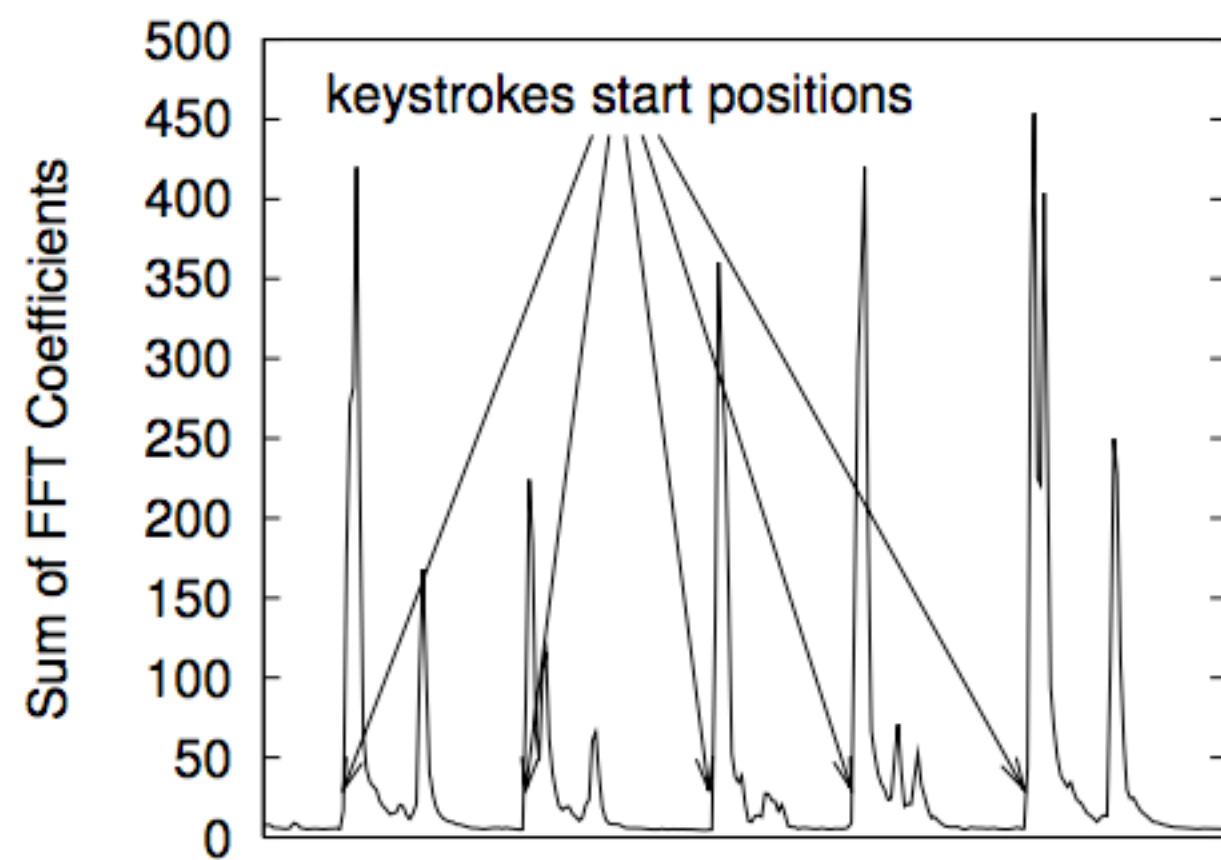
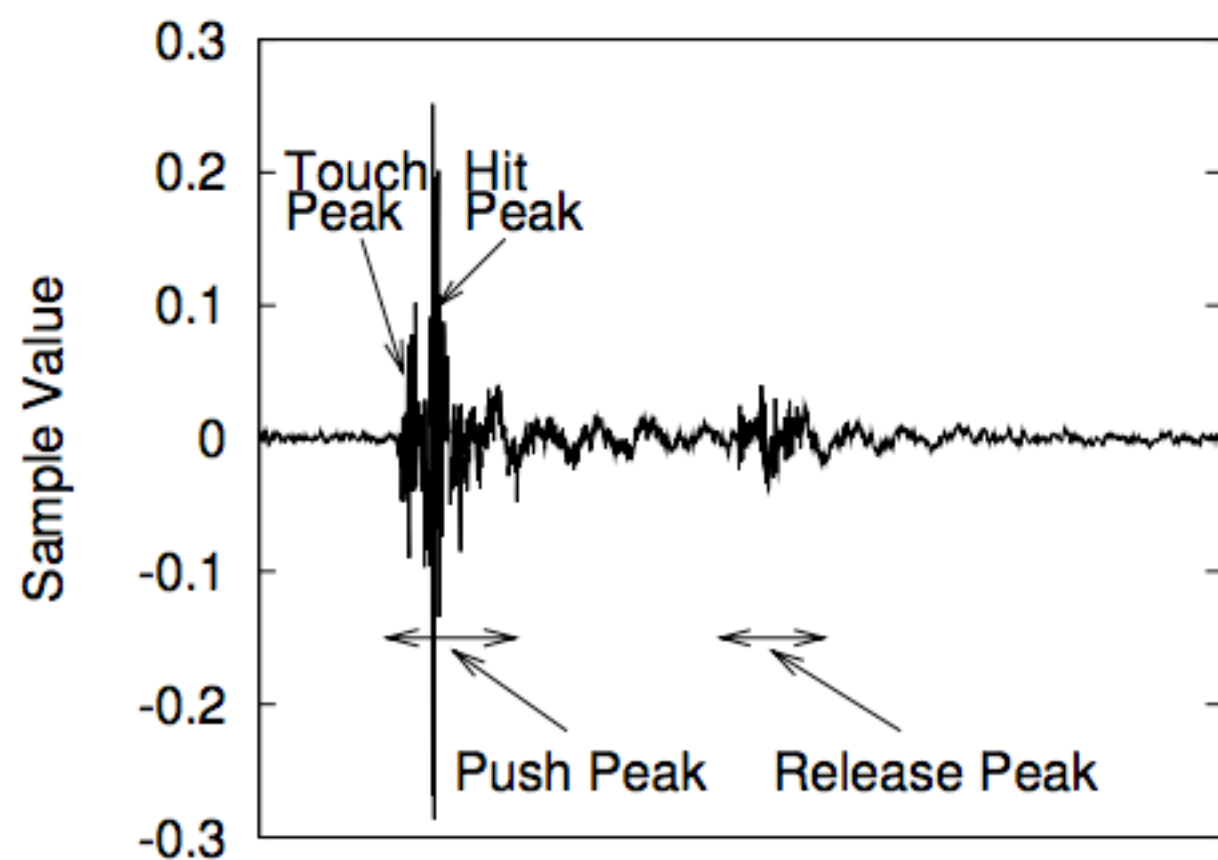
- What are the challenges?
 - Assumptions = Deviation from real world
 - Bias
 - Keeping the model simple

CHALLENGES



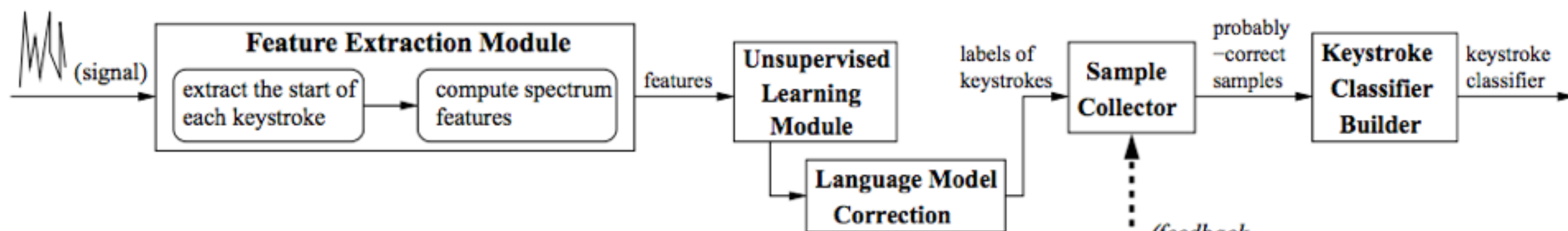
3. ACOUSTIC KEYLOGGER

SAMPLE DATA

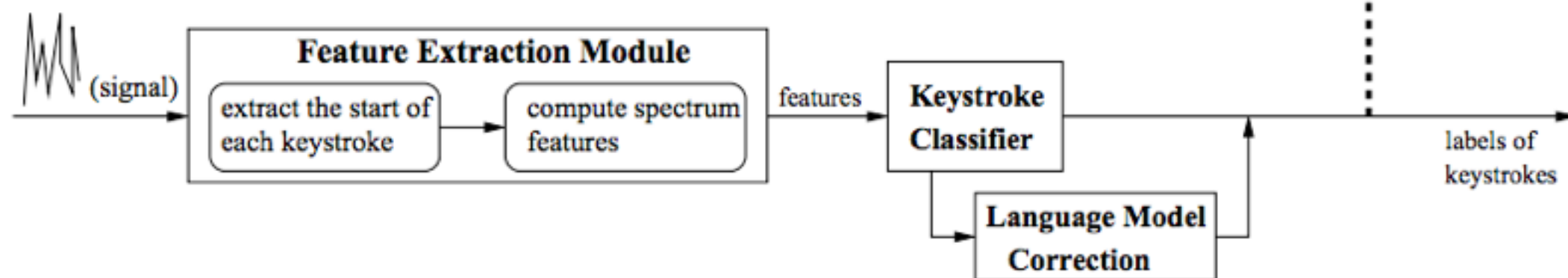


DEMO!

UNSUPERVISED METHOD



(a) Training Phase: Build keystroke classifier using unsupervised learning



(b) Recognition Phase: Recognize keystrokes using the classifier from (a).

ACOUSTIC KEYLOGGER

- Future direction
 - Dealing with noise, before training
 - Signal amplification
 - Introducing enough independent variables
 - Preparing a more comprehensive training dataset

REAL WORLD ATTACK VECTORS

iPhone in hand or on table

Recording over conference call

Skype

Hidden microphone/contact microphone

Directional microphone in public space

THANKS!

Scott Stevenson

@scottastevenson

scottstevenson88@gmail.com

Navid Shekoufa

@n6599

shekoufa.navid@gmail.com

meetup.ndev.co

slack.ndev.co