Machine Learning for Signal Processing

Lecture 1: Introduction
Representing sound and images

Class 1. 28 Aug 2018
Instructor: Bhiksha Raj
What is a signal

- A mechanism for conveying information
  - Semaphores, gestures, traffic lights...

- In Electrical Engineering: currents, voltages

- Digital signals: Ordered collections of numbers that convey information
  - from a source to a destination
  - about a real world phenomenon
    - Sounds, images
Signal Examples: Audio

• A sequence of numbers
  – \([n_1 \, n_2 \, n_3 \, n_4 \, \ldots]\)
  – The order in which the numbers occur is important
    • Ordered
    • In this case, a *time series*
  – Represent a perceivable sound
Example: Images

• A rectangular arrangement (matrix) of numbers
  – Or sets of numbers (for color images)
• Each pixel represents a visual representation of one of these numbers
  – 0 is minimum (black), 1 is maximum (white)
  – Position / order is important

Pixel = 0.5
Example: Biosignals

- MRI: “k-space” $\rightarrow$ 3D Fourier transform
  - Invert to get image
- EEG: Many channels of brain electrical activity
- ECG: Cardiac activity
- OCT, Ultrasound, Echo cardiogram: Echo-based imaging
- Others..
Financial Data

- Stocks, options, other derivatives
- Analyze trends and make predictions
- Special Issues on Signal Processing Methods in Finance and Electronic Trading from various journals
Many others

- Network data..
- Weather..
- Any stochastic time series
- Etc.
What is Signal Processing

• Acquisition, Analysis, Interpretation, and Manipulation of signals.
  – Acquisition:
    • Sampling, sensing
  – Analysis:
    • Decomposition: Separating signals into basic “building” blocks
  – Manipulation:
    • Denoising
    • Coding: GSM, Jpeg, Mpeg, Ogg Vorbis
  – Interpretation:
    • Detection: Radars, Sonars
    • Pattern matching: Biometrics, Iris recognition, finger print recognition
    • Prediction: Financial prediction, speech coding, etc.
  – Etc.
• Boundaries between these categories of operations are fuzzy
The Tasks in a typical Signal Processing Paradigm

• Capture: Recovery, enhancement
• Channel: Coding-decoding, compression-decompression, storage
• Regression: Prediction, classification
What is Machine Learning

• The science that deals with the development of algorithms that can learn from data
  – Learning the structure of data
    • Feature extraction
  – Learning patterns in data
    • Automatic text categorization; Market basket analysis
  – Learning to classify between different kinds of data
    • Is that picture a flower or not?
  – Learning to predict data
    • Weather prediction, movie recommendation

• Statistical analysis and pattern recognition when performed by a computer scientist..
MLSP

• Application of Machine Learning techniques to the analysis of signals

• Can be applied to each component of the chain

(sensor)

Signal Capture ➔ Channel ➔ Feature Extraction ➔ Modeling/Regression

Can be applied to each component of the chain
MLSP

- Application of Machine Learning techniques to the analysis of signals

- *Can be applied to each component of the chain*

- Sensing
  - Compressed sensing, dictionary based representations

- Denoising
  - ICA, filtering, separation
MLSP

• Application of Machine Learning techniques to the analysis of signals

• Can be applied to each component of the chain

• Channel: Compression, coding
MLSP

- Application of Machine Learning techniques to the analysis of signals

- Can be applied to each component of the chain

- Feature Extraction:
  - Dimensionality reduction
    - Linear models, non-linear models

Diagram:

```
sensor

| Signal Capture | Channel | Feature Extraction | Modeling/Regression |
```

11-755/18-797 14
MLSP

• Application of Machine Learning techniques to the analysis of signals

• Can be applied to each component of the chain

• Classification, Modelling and Interpretation, Prediction
In this course

• The four “aspects” of MLSP:

  – **Representation**: How best to represent signals for effective downstream or upstream processing
  
  – **Modeling**: How to *model* the systematic and statistical characteristics of the signal
  
  – **Classification**: How do we assign a class to the data?
  
  – **Prediction**: How do we predict new or unseen values or attributes of the data
What we will cover

• **Representations**: Algebraic methods for extracting information from signals
  – Deterministic representations
  – Data-driven characterization
    • PCA
    • ICA
    • NMF
    • Factor Analysis
    • LGMs
What we will cover

• **Representations/Modelling**: Learning-based approaches for modeling data
  – Dictionary representations
  – Sparse estimation
    • Sparse and over-complete characterization, Compressed sensing
  – Regression
  – Neural networks

• **Modelling**: Latent variable characterization
  – Clustering, K-means
  – Expectation Maximization
  – Probabilistic Latent Component Analysis
What we will cover

• **Modeling/Prediction:** Time Series Models
  – Markov models and Hidden Markov models
  – Linear and non-linear dynamical systems
    • Kalman filters, particle filtering

• **Classification and Prediction:**
  – Binary classification. Meta-classifiers
  – Neural networks

• **Wish list: Additional topics**
  – Privacy in signal processing
  – Extreme value theory
  – Dependence and significance
Recommended Background

- **DSP**
  - Fourier transforms, linear systems, basic statistical signal processing

- **Linear Algebra**
  - Definitions, vectors, matrices, operations, properties

- **Probability**
  - Basics: what is an random variable, probability distributions, functions of a random variable

- **Machine learning**
  - Learning, modelling and classification techniques
Guest Lectures

- TBD
Schedule of Other Lectures

• Tentative Schedule on Website
• [http://mlsp.cs.cmu.edu/courses/fall2018](http://mlsp.cs.cmu.edu/courses/fall2018)
  – To be updated before Thursday, currently buggy
Grading

• Mini quizzes : 25%
  – Ten multiple-choice questions on the topics of the week
  – Weekly
  – Will be open on Friday, closed on Saturday night

• Homework assignments : 50%
  – Mini projects
  – Will be assigned during course
  – Expect four
  – You will not catch up if you slack on any homework
    • Those who didn’t slack will also do the next homework

• Final project: 25%
  – Will be assigned early in course
  – Dec 6 (approx): Poster presentation for all projects, with demos (if possible)
    • Partially graded by visitors to the poster
Instructor and TA

• Instructor: Prof. Bhiksha Raj
  – Room 6705 Hillman Building
  – bhiksha@cs.cmu.edu
  – 412 268 9826

• TAs:
  – Abelino Jimenez (abelinoj@andrew)
  – Benjamin Elizalde (benmael@andrew)
  – Yandong Wen (yandongw@andrew)
  – Kigali: Samuel Ishimwe (sishimwe@Africa.cmu.edu)
  – SV: ??

• Office Hours:
  – Instructor: Thursday, 1-2.30; I also have an open-door policy
  – TAs: TBD
Additional Administrivia

• Website:
  – Lecture material will be posted on the day of each class on the website
  – Reading material and pointers to additional information will be on the website

• We will use Piazza
  – Expect an invite to join 11-755/18-797

• Mailing list: Information will be posted
Continuing..

• Story so far:
  – What is a signal
  – Some types of signals
  – What is SP
  – What is ML
    • And where does it apply in the SP chain
• Continuing – some additional concepts..
  – More on signals
  – More on what we do with signals
    • Representation, Regression, classification, prediction
  – And how
    • Supervision
More on Signals

• Principles of signal *capture* and what the numbers mean

• Explained through examples
  – Sound, images
    • Signals where the purpose of signal capture is to recreate stimulus
    • Signals we emphasize a bit in course
    • But also because easily interpretable principles that extend to all signals
  – Also MRI
    • Illustrates capture in *transform* domain
E.g. Audio Signals

- A typical digital audio signal
  - It’s a sequence of numbers
  - Must represent a quantity that enables near-perfect recreation of sound stimulus
The sound stimulus

• Any sound is a pressure wave: alternating highs and lows of air pressure moving through the air
• When we speak, we produce these pressure waves
  – Essentially by producing puff after puff of air
  – Any sound producing mechanism actually produces pressure waves
• These pressure waves move the eardrum
  – Highs push it in, lows suck it out
  – We sense these motions of our eardrum as “sound”
SOUND PERCEPTION
SOUND PERCEPTION
Storing pressure waves on a computer

• The pressure wave moves a diaphragm
  – On the microphone

• The motion of the diaphragm is converted to continuous variations of an electrical signal
  – Many ways to do this

• A “sampler” samples the continuous signal at regular intervals of time and stores the numbers
Are these numbers sound?

- How do we even know that the numbers we store on the computer have anything to do with the recorded sound really?
  - Recreate the sense of sound
- The numbers are used to control the levels of an electrical signal
Are these numbers sound?

- How do we even know that the numbers we store on the computer have anything to do with the recorded sound really?
  - Recreate the sense of sound
- The numbers are used to control the levels of an electrical signal
- The electrical signal moves a diaphragm back and forth to produce a pressure wave
  - That we sense as sound
How many samples a second

• Convenient to think of sound in terms of sinusoids with frequency

• Sounds may be modelled as the sum of many sinusoids of different frequencies
  – Frequency is a physically motivated unit
  – Each hair cell in our inner ear is tuned to specific frequency

• Any sound has many frequency components
  – We can hear frequencies up to 16000Hz
    • Frequency components above 16000Hz can be heard by children and some young adults
    • Nearly nobody can hear over 20000Hz.
Signal representation - Sampling

- *Sampling frequency* (or *sampling rate*) refers to the number of samples taken a second

- Sampling rate is measured in Hz
  - We need a sample rate **twice as high** as the highest frequency we want to represent (Nyquist freq)

- For our ears this means a sample rate of at least 40kHz
  - Because we hear up to 20kHz
Aliasing

- Low sample rates result in *aliasing*
  - High frequencies are misrepresented
  - Frequency $f_1$ will become $(\text{sample rate} - f_1)$
  - In video also when you see wheels go backwards
Aliasing examples

Sinusoid sweeping from 0Hz to 20kHz

44.1kHz SR, is ok
22kHz SR, aliasing!
11kHz SR, double aliasing!

On real sounds
at 44kHz
at 22kHz
at 11kHz
at 5kHz
at 4kHz
at 3kHz

On images

On video
Aliasing Examples

Sinusoid sweeping from 0Hz to 20kHz

11kHz SR, double aliasing!
44.1kHz SR, is ok
22kHz SR, aliasing!
11kHz SR, double aliasing!

On real sounds
at 44kHz
at 22kHz
at 11kHz
at 5kHz
at 4kHz
at 3kHz

On images

On video
Avoiding Aliasing

- Solution: *Filter the signal* before sampling it
  - Cut off all frequencies above sampling.frequency/2
  - E.g., to sample at 44.1Khz, filter the signal to eliminate all frequencies above 22050 Hz

- Will only lose information, but not distort existing information
Problem 2: Resolution

• Sound is the outcome of a continuous range of variations
  – The pressure wave can take any value (within limits)

• A computer has finite resolution
  – Numbers can only be stored to finite resolution
  – E.g. a 16-bit number can store only 65536 values, while a 4-bit number can store only 16 unique values

• Low-resolution storage results in loss of information
Storing the signal on a computer

- The original signal
- 8 bit quantization
- 3 bit quantization
- 2 bit quantization
- 1 bit quantization
Tom Sullivan Says his Name

- 16 bit sampling
- 5 bit sampling
- 4 bit sampling
- 3 bit sampling
- 1 bit sampling
A Schubert Piece

- 16 bit sampling
- 5 bit sampling
- 4 bit sampling
- 3 bit sampling
- 1 bit sampling
Lessons (for any signal)

• *Transduce* signal in meaningful manner
  – For sound and images, must be able to recreate original stimulus from signal

• Sample fast enough to capture highest frequency variations

• Store with sufficient resolution

• For audio
  – Common sample rates
    • For speech 8kHz to 16kHz
    • For music 32kHz to 44.1kHz
    • Pro-equipment 96kHz
  – Common bit resolution
    • 12-bit equivalent for speech
    • 16 bits for high-fidelity speech
    • 24 bits for music
Images
Images
The Eye

Vitreous humor  Lens  Aqueous humor  Cornea  Air

Retina

Image  Object

http://www.brad.ac.uk/acad/lifesci/optometry/resources/modules/stage1/pvp1/Retina.html
Rods and Cones

- Separate Systems
- Rods
  - Fast
  - Sensitive
  - Grey scale
  - predominate in the periphery
- Cones
  - Slow
  - Not so sensitive
  - Fovea / Macula
  - COLOR!

The density of cones is highest at the fovea
  – The region immediately surrounding the fovea is the macula
    • The most important part of your eye: damage == blindness
• Peripheral vision is almost entirely black and white
• Eagles are bifoveate
• Dogs and cats have no fovea, instead they have an elongated slit
Three Types of Cones (trichromatic vision)
Trichromatic Vision

• So-called “blue” light sensors respond to an entire range of frequencies
  – Including in the so-called “green” and “red” regions

• The difference in response of “green” and “red” sensors is small
  – Varies from person to person
    • Each person really sees the world in a different color
  – If the two curves get too close, we have color blindness
    • Ideally traffic lights should be red and blue
White Light
Response to White Light
Response to White Light
Response to Sparse Light
Response to Sparse Light
Response to Sparse Light

Many combinations of wavelength can produce the same perception of colour.
Several types of animals are *tetrachromatic* (including at least one human)

Estrildid finches
Digital Capture of Images

- Lens projects image on sensor
  - Typically CCD or CMOS
- Sensor comprises sensing elements of 3 colors
  - Different strategies for arrangement of color sensors
- Limited number of sensing elements
  - 200-600 ppi
  - The camera generally includes an anti-aliasing filter to eliminate aliasing in the image
Representing Images

• Utilize trichromatic nature of human vision
  – Trigger the three cone types to produce a sensation approximating desired color
    • A *tetrachromatic* animal would be very confused by our computer images

• The three “chosen” colors are red (650nm), green (510nm) and blue (475nm)

• Can still only represent a small fraction of the 10 million colors that humans can sense
Computer Images: Grey Scale

Signal: Each stored number represents a single pixel

R = G = B. Only a single number need be stored per pixel

Picture Element (PIXEL)
Position & gray value (scalar)
What we see

What the computer “sees”
Color Images

Signal: Each triad of stored numbers represents a single pixel

Picture Element (PIXEL)
Position & color value (red, green, blue)
RGB Representation

original

11-755/18-797

R

G

B
The CMYK color space

Represent colors in terms of cyan, magenta, and yellow
- The “K” stands for “Key”, not “black”
CMYK is a *subtractive* representation

- RGB is based on *composition*, i.e. it is an additive representation
  - Adding equal parts of red, green and blue creates white
- What happens when you mix red, green and blue paint?
  - Clue – paint colouring is subtractive.
- CMYK is based on *masking*, i.e. it is subtractive
  - The base is white
  - Masking it with equal parts of C, M and Y creates Black
  - Masking it with C and Y creates Green
    - Yellow masks blue
  - Masking it with M and Y creates Red
    - Magenta masks green
  - Masking it with M and C creates Blue
    - Cyan masks green
  - Designed specifically for *printing*
    - As opposed to rendering
An Interesting Aside

- Paints create subtractive coloring
  - Each paint masks out some colours
  - Mixing paint subtracts combinations of colors
  - Paintings represent subtractive colour masks

- In the 1880s Georges-Pierre Seurat pioneered an *additive-colour* technique for painting based on “pointilism”
  - How do you think he did it?
Quantization and Saturation

- Captured images are typically quantized to 8 bits
- 8-bits is not very much < 1000:1
- Humans can easily accept 100,000:1
- And most cameras will give you only 6-bits anyway...
  - Truth in advertising!
Processing Colour Images

• Typically work only on the Grey Scale image
  – Decode image from whatever representation to RGB
  – GS = R + G + B

• For specific algorithms that deal with colour, individual colours may be maintained
  – Or any linear combination that makes sense may be maintained.
Signals..

• Speech and Images are examples of signals where the digitized signal is a facsimile of the stimulus to be represented
  – Many other signals of this kind, including bio-signals, network traffic, etc.

• Next up: a signal where the digitized signal is not a direct facsimile of the data to be represented
  – Signal captured in a *transform* domain
Magnetic Resonance Imaging

- Attempts to image *interior structure* of soft tissue
- Does so by imposing a magnetic field and measuring resonance of protons (Hydrogen atoms)
Cross-section of a body

- Image changes left to right, top to bottom at different rates at different locations
  - Different tissue densities...
  - ... which show up as a range of “spatial frequencies”
MRI

- Takes slice-wise measurements in the Fourier domain
- A single “gradient field” derives response from a single “spatial frequency” component
  - Which can be measured
- Sequence of gradient fields derive resonant response of different spatial frequencies of tissue slice
  - Effectively a 2D Fourier transform
- Must invert transform to create image
- “Join” slices for full 3-D reconstruction
What we do with signals

- Have seen examples of signals and caveats of signal capture
- Next: Machine Learning challenges in dealing with the data
• **Signals** can be *decomposed* into combinations of building blocks
  – Different signals of any category composed as different combinations of the same building blocks
  – Knowing the composing combination informs us about the properties of the signal
    • But requires knowing the building blocks
    • Using the wrong building blocks will give us imprecise or meaningless conclusions

• **ML challenge**: Find building blocks from analysis of signals
  – Mathematically: \( S = f(B, W) \), find \( B \) and \( W \) from \( S \)
  – \( S = \) signal, \( B = \) building blocks, \( W = \) combination parameters, \( f = \) combination function
Modelling

- Signals are produced by *processes*
  - Which are generally partially or fully unknown
- Knowledge of the process is often crucial for additional processing
  - Control, prediction, analysis
- **ML challenge**: Characterizing the process underlying the signal
  - Characterization through statistical properties of the signal
  - Characterization through an abstract parametric model
Classification

• Signals may arise from different classes of stimuli/processes
• Often needed to identify underlying process/stimulus
• **ML challenge**: Identify underlying “class” of the signal
• Signals can be analyzed to make predictions about the future of the signal or the underlying process

• **ML challenge**: How to make the “best” predictions
Supervision

• Learning representations and modeling are often preliminary steps to classification and prediction

• Can be performed *without* reference to the actual classification/prediction task
  – *Unsupervised* learning

• Can be explicitly optimized
Supervision

- Task: Detect if it’s a face
- Unsupervised representation: characterize edges, gradation
  - Does not specifically help with problem
- Supervised representation: characterize nose-like features, eyebrow-like features, mouth-like features...
  - Better suited to detect faces
Primary tools of the trade..

• Linear algebra
  – Some calculus

• Optimization

• Probability...
Machine learning problems often require finding parameters/values that “optimize” an objective.

Typical objectives:
- Error of constructing a signal
- Accuracy of predicting future
- Error in classifying signal

Problem: Given only variation of objective w.r.t. parameters of algorithm, find the optimal set of parameters.
Optimization: Formulation

• In the majority of machine learning task, a set of samples is provided \( z_1, z_2, \ldots, z_n \)

• Supervised learning
  \[
  z = (x, y) \\
  h \text{ is predictor function } h : X \to Y \\
  \text{minimize } f(h;(x,y)) = \text{loss function}(h(x),y)
  \]

• Unsupervised learning (k-mean clustering)
  \[
  z = x \in \mathbb{R}^d \\
  h = (\mu_1, \ldots, \mu_k) \in \mathbb{R}^{d \times k}, \text{ which corresponds to cluster centers} \\
  \text{minimize } f((\mu_1, \ldots, \mu_k);x) = \min_j \left\| \mu_j - x \right\|^2
  \]
Next Class..

• Review of linear algebra..