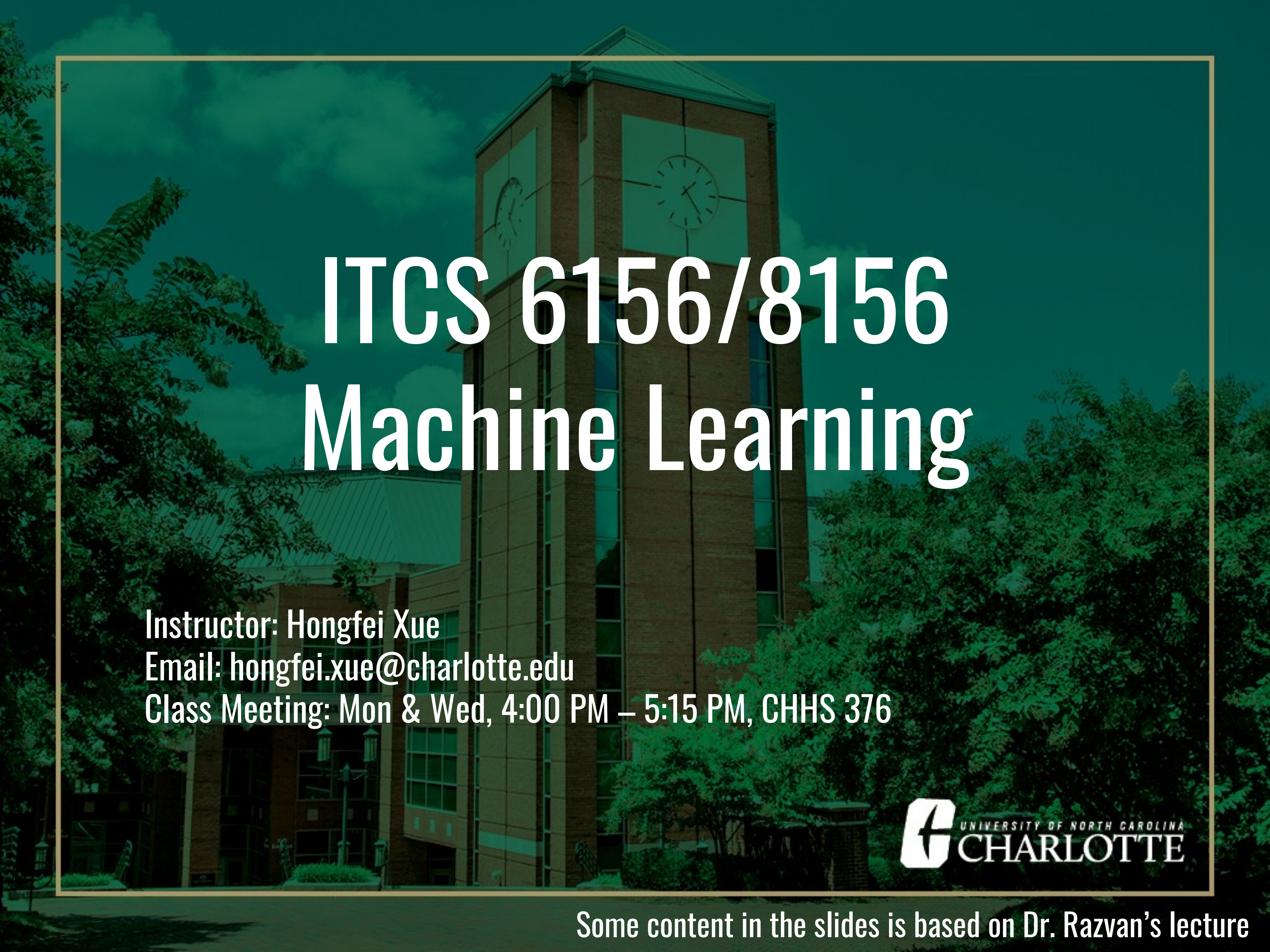


ITCS 6156/8156
Machine Learning

Instructor: Hongfei Xue
Email: hongfei.xue@charlotte.edu
Class Meeting: Mon & Wed, 4:00 PM – 5:15 PM, CHHS 376

UNIVERSITY OF NORTH CAROLINA
CHARLOTTE

Some content in the slides is based on Dr. Razvan’s lecture

How to Automate Solutions to Computational Problems

- Spam email classification:
  - Binary classification of emails: Spam vs. Ham (Legitimate message)

- Expert Systems approach (Rule-based)
  - A group of experts write rules determining whether an email is spam or not.
  - A programmer implement the rules into computer code

- Example rules:
  - Classify the email as spam in “Money” appears in the text.
  - What if the email is sent by your parents?

How to Automate Solutions to Computational Problems

- Cons of **Expert Systems** approach (Rule-based)
  - **Cognitively demanding**: Difficult for humans to reason with many useful but imprecise features that are indicative (signals) of spam or not spam:

    - Words, phrases, images, meta-data, time series, …

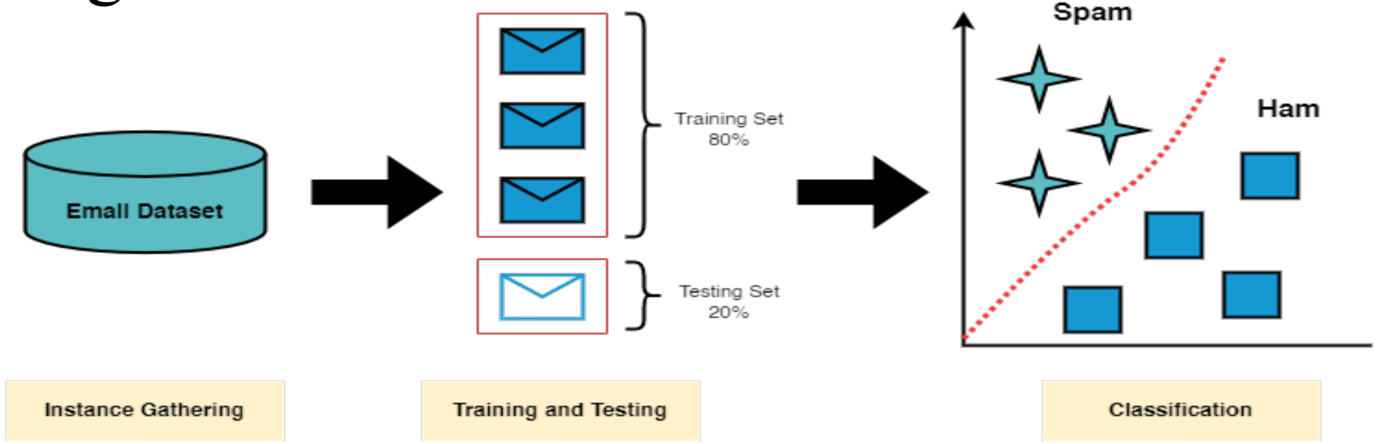
    - Need to combine a large number of signals, figure out their relative importance in determining spam vs. ham label.
  - **Brittle**: Always going to miss some usef features or patterns

    - Spam filtering is adversarial, new features need be added over time.

Expert (Rule-based) Systems

Why Machine Learning?

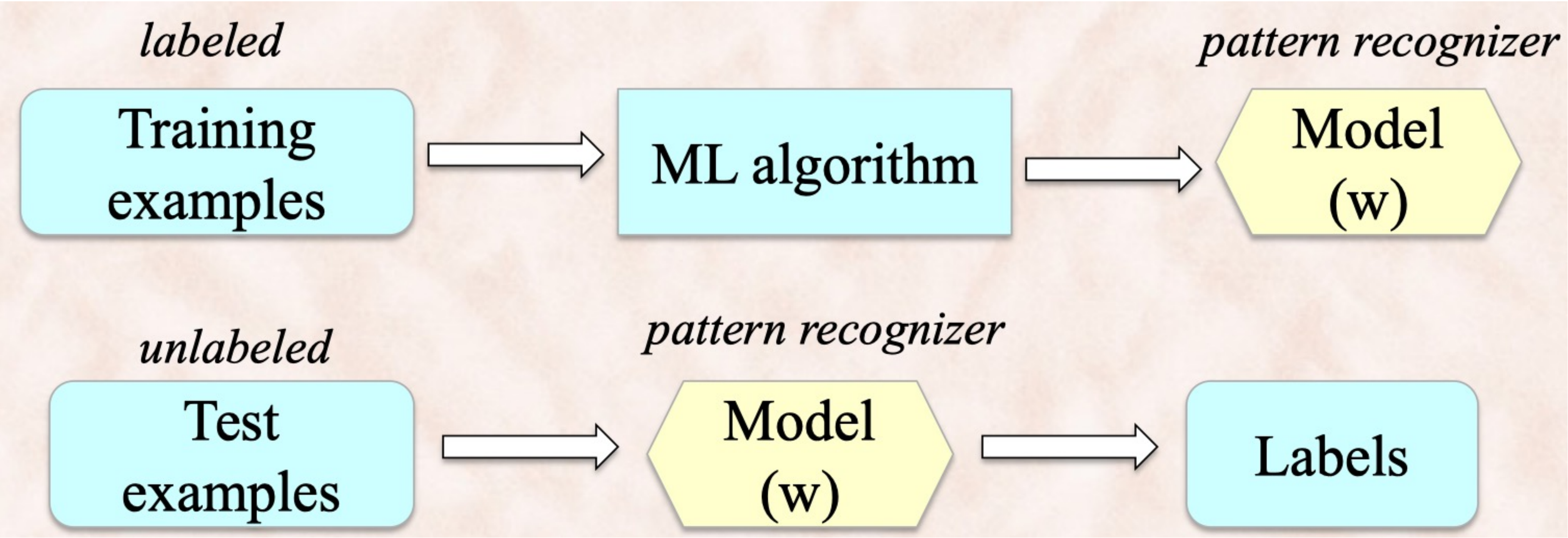
- Machine Learning algorithms can automatically learn the weights to combine features.



- A typical Machine Learning (ML) approach:
  1. Acquire a large enough dataset of labeled examples:
     - Each email is an instance, the label is spam (+1) vs. not spam (-1).
  2. Represent emails as feature vectors:
     - Each feature has a weight, the sign of the weighted sum of features should match the label.
       - Traditional ML: Engineer the features.
       - Deep ML: Learn the features
  3. Learn the weights so that the model (weighted combination of features) does well on labeled examples.

What's Machine Learning?

- **Machine Learning** is to construct computer programs that learn from experience to perform well on a given task.
  - **Supervised Learning**: discover patterns from labeled examples that enable predictions on (previously unseen) unlabeled examples.



Questions?