CMSC 678
Introduction to Machine Learning
Spring 2018

https://www.csee.umbc.edu/courses/graduate/678/spring18/

Some slides adapted from Hamed Pirsiavash
Frank Ferraro

ITE 358
ferraro@umbc.edu
Monday: 3:45-4:30
Tuesday: 11-11:30
by appointment

Natural language processing:
- Semantics

Vision & language processing

Generative & neural modeling

Learning with low-to-no supervision
TA: Vamshi Nagabandi

Location TBA

nvamshi1@umbc.edu

Machine learning

Wednesday 1-2

Thursday 2:30-3:30

Data analytics
The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS  DEC. 14, 2016
Les talibans mènent des attaques-suicides dans Kaboul

Près de 17 personnes, dont un Français et un Italien, ont été tuées dans une série d'attaques revendiquées par les talibans. C'est l'attaque la plus meurtrière depuis janvier.

Air France s'attend à une perte historique pour l'exercice 2009-2010

La direction annonce que la compagnie va perdre 1,3 milliard d'euros sur l'exercice qui sera clos fin mars.
Hi all,
We wanted to invite you to join us for an early Thanksgiving on November 22nd, beginning around 2PM. Please bring your favorite dish! RSVP by next week.

Dave

Hi team,

The server appears to be dropping about 10% of requests (see attached dashboards). There hasn't been a new release since last night, so I'm not sure what's going on. Is anyone looking into this?

...
"This country made a promise to every hardworking American: Social Security will be there when you need it. America must honor its promise and that means no cuts to the Social Security budget."
— Sen. Elizabeth Warren

Americans Demand Full Funding of the Social Security

Last week, Senators Bernie Sanders (I-VT), Bob Casey (D-PA), and Eliz...

MEDIUM.COM

2.9K 566 410

"The Democrats have taken all talk of the DACA kids and the DREAMers off the table for the next round of budget talks in 2-1/2 weeks to prevent a government shutdown. So not only did Schumer remove the $25 billion he was gonna offer Trump for the wall, they're removing all of their demands for the DREAMers. It's a total backtrack. They've caved again to Trump."

From Rush's Famous Quotes, updated live throughout the show at RushLimbaugh.com
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)

This is *not* a survey course.

We will go deep into the topics.
Building the Next New York Times Recommendation Engine

When artificial intelligence makes a picture worth way more than a thousand words

Facebook AI Still Can’t Do Things Even A Baby Has Mastered

Shifting toward the Knowledge Economy

Google's AI can translate language pairs it has never seen
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)

Learn different ways that success and progress can be measured in ML
Deep Learning

What society thinks I do
What my friends think I do
What other computer scientists think I do

What mathematicians think I do
What I think I do
What I actually do

from theano import *
keras
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)
Learn different ways that success and progress can be measured in ML
Relate to statistics, AI [671], and specialized areas (e.g., NLP [673] and CV [691])
Implement ML programs
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)
Learn different ways that success and progress can be measured in ML
Relate to statistics, AI [671], and specialized areas (e.g., NLP [673] and CV [691])
Implement ML programs

Assignments will require your own implementation.
Course Goals

Be introduced to some of the core problems and solutions of ML (big picture)
Learn different ways that success and progress can be measured in ML
Relate to statistics, AI [671], and specialized areas (e.g., NLP [673] and CV [691])
Implement ML programs
Read and analyze research papers
Practice your (written) communication skills
Administrivia
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>

Each component is \( \max(\text{micro-average}, \text{macro-average}) \)
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>

max(micro-average, macro-average)

- 65/90
- 95/100
- 95/110
- 100/110
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>

\[
\text{max(\text{micro-average}, \text{macro-average})}
\]

\[
\text{microaverage} = \frac{65 + 95 + 95 + 100}{90 + 100 + 110 + 110} \approx 86.59\%
\]
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>

\[
\text{max(micro-average, macro-average)}
\]

\[
\text{microaverage} = \frac{65 + 95 + 95 + 100}{90 + 100 + 110 + 110} \approx 86.59\%
\]

\[
\text{macroaverage} = \frac{1}{4} \left( \frac{65}{90} + \frac{95}{100} + \frac{95}{110} + \frac{100}{110} \right) \approx 86.12\%
\]
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Course Project</td>
<td>40%</td>
</tr>
<tr>
<td>Two Exams</td>
<td>20%</td>
</tr>
</tbody>
</table>

\[
\text{max}(\text{micro-average}, \text{macro-average})
\]

<table>
<thead>
<tr>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>65/90</td>
<td>95/100</td>
<td>95/110</td>
<td>100/110</td>
</tr>
</tbody>
</table>

\[
\text{microaverage} = \frac{65 + 95 + 95 + 100}{90 + 100 + 110 + 110} \approx 86.59\%
\]

\[
\text{macroaverage} = \frac{1}{4} \left( \frac{65}{90} + \frac{95}{100} + \frac{95}{110} + \frac{100}{110} \right) \approx 86.12\% \]
# Final Grades

<table>
<thead>
<tr>
<th>If you get ≥</th>
<th>You get at least a/an</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>A-</td>
</tr>
<tr>
<td>80</td>
<td>B-</td>
</tr>
<tr>
<td>70</td>
<td>C-</td>
</tr>
<tr>
<td>65</td>
<td>D</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>
Introduction to Machine Learning

Spring 2018 — CMSC 678

Announcements

- 1/29/18 Assignment 1 is available. It's due 2/7/18 by 11:59 AM.
- 1/29/18 The syllabus is available.
- 1/29/18 See the Piazza discussion board for class announcements and Q&A.

Who, What, When, and Where

Check out the syllabus for all this information, including policies on academic honesty, accommodations, and late assignments.

Meeting Times
Performing Arts and Humanities, 132
Monday & Wednesday, 2:30pm - 3:45pm

Instructor
Frank Ferraro
ferraro [at] umbc [dot] edu
ITE 358
Monday 3:45 - 4:30
Tuesday 11:00 - 11:30
by appointment

TA
Vamshi Nagabandi
nvamsh1 [at] umbc [dot] edu
Location TBA
Wednesday 1:00 - 2:00
Thursday 2:30 - 3:30

Topics
The topics covered will include, but are not limited to:
- perceptrons
- regression (linear, logistic, and non-linear)
- spectral, clustering, and dimensionality reduction techniques
- support vector machines and kernel methods
- neural networks, including deep learning, recurrent neural networks
- Bayesian networks and probabilistic graphical models
- clustering
- evaluation methodologies and experiment design.

Goals
After taking this course, you will
- be introduced to some of the core problems and solutions in modern data analysis and machine learning
- develop the skills needed to apply machine learning and data analysis methods to real-world problems
Submitting Your Work

https://www.csee.umbc.edu/courses/graduate/678/spring18/submit
Running the Assignments

A "standard" x86-64 Linux machine, like gl

A passable amount of memory (2GB-4GB)

Modern but not necessarily cutting edge software

Don’t assume a GPU (if you want to write CUDA yourself, talk to me)

If in doubt, ask first
Running the Project

An x86-64 Linux machine

Memory and hardware constraints lifted (somewhat)

If in doubt, ask first
Programming Languages for Assignments

Use the tools you feel comfortable with

Python+numpy, C, C++, Java, Matlab, ...: OK (straight Python may not cut it)

Libraries: Generally OK, as long as you don’t use their implementation of what you need to implement

Math accelerators (blas, numpy, etc.): OK

If in doubt, ask first
Programming Languages for the Project

Use the tools you feel comfortable with

Python+numpy, C, C++, Java, Matlab, ...: OK (straight Python may not cut it)

Libraries: Use what you want

Math accelerators (blas, numpy, etc.): OK
Online Discussions

google.com/piazza.com/umbc/spring2018/cmsc678
## Important Dates

<table>
<thead>
<tr>
<th>Date</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, 2/7</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>Monday, 3/5</td>
<td>Assignment 2</td>
</tr>
<tr>
<td>Monday, 3/12</td>
<td>Project Proposal</td>
</tr>
<tr>
<td>Wednesday, 3/14</td>
<td>Exam 1 (In-class)</td>
</tr>
<tr>
<td>Monday, 4/2</td>
<td>Assignment 3</td>
</tr>
<tr>
<td>Monday, 4/9</td>
<td>Project Update</td>
</tr>
<tr>
<td>Monday, 5/14</td>
<td>Assignment 4</td>
</tr>
<tr>
<td>Friday, 5/18</td>
<td>Exam 2 (Final exam block)</td>
</tr>
<tr>
<td>Wednesday, 5/23</td>
<td>Course Project</td>
</tr>
</tbody>
</table>

All items due **11:59 AM** UMBC time (unless specified otherwise)
Late Policy

Everyone has a budget of 10 late days
Late Policy

Everyone has a budget of 10 late days

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked
Late Policy

Everyone has a budget of 10 late days.

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked.

If you don’t have any left: still turn assignments in. They could count in your favor in borderline cases.
Late Policy

Everyone has a budget of 10 late days

Use them as needed throughout the course
They’re meant for personal reasons and emergencies

Do not procrastinate
Late Policy

Everyone has a budget of 10 late days

Contact me privately if an extended absence will occur

You must know how many you’ve used
Resource #1: ESL

“Elements of Statistical Learning”
Hastie, Tibshirani, Friedman


Full book: 

Official: Recommended
Resource #2: ITILA

“Information Theory, Inference and Learning Algorithms”
MacKay

http://www.inference.org.uk/mackay/itprnn/ps/

Full book:

Official: Recommended
Resource #3: UML

“Understanding Machine Learning: From Theory to Algorithms”
Shalev-Shwartz, Ben-David

http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/


Official: Recommended
Resource #4: CIML

“A Course in Machine Learning”, v0.99
Hal Daumé III

http://ciml.info/

Full book:
http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf

Unofficial
Resources #5... ∞

Peer-reviewed articles (journals, conferences & workshops)

ICML

NIPS

JMLR

CVF
Who should take this course?

Is this the right course for you?
  good math and programming background?
  diligent and determined?
  willing to implement & write up your results?

Unsure? Let’s talk after class

(thank you to everyone who filled out the survey! :) )
https://goo.gl/forms/yqVH8QnwzggpRQJr1
Why do we care about math?!

Calculus and linear algebra
  Techniques for finding maxima/minima of functions
  Convenient language for high dimensional data analysis

Probability
  The study of the outcomes of repeated experiments
  The study of the plausibility of some event

Statistics
  The analysis and interpretation of data
Course Announcement 1: Assignment 1

Due Wednesday, 2/7 (~9 days)

Math & programming review

Discuss with others, but write, implement and complete on your own
What does it mean to learn?

Chris has just begun taking a machine learning course

Pat, the instructor has to ascertain if Chris has “learned” the topics covered, at the end of the course

What is a “reasonable” exam?

(Bad) Choice 1: History of pottery
    Chris’ s performance is not indicative of what was learned in ML
(Bad) Choice 2: Questions answered during lectures
    Open book?

A good test should test ability to answer “related” but “new” questions on the exam

Generalization
Machine Learning Framework: Learning

instance 1

instance 2

instance 3

instance 4

Machine Learning Predictor
Machine Learning Framework: Learning

instances are typically examined independently

Machine Learning Predictor

Extra-knowledge
Machine Learning Framework: Learning

instances are typically examined independently
Machine Learning Framework: Learning

instances are typically examined independently
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
scoring model

\[ \text{score}_{\theta}(X) \]

objective

\[ F(\theta) \]
scoring model \( \text{score}_{\theta}(X) \)

objective \( F(\theta) \)

(implicitly) dependent on the observed data \( X \)
Gradient Ascent

$$\arg \max_{\theta} \quad F(\theta)$$
Gradient Ascent

\[ \arg \max_{\theta} F (\theta) \]
Gradient Ascent

\[
\arg \max_{\theta} F(\theta)
\]
Gradient Ascent

$$\arg \max_{\theta} \ F (\theta)$$
Underfitting and overfitting

Images courtesy Hamed Pirsiavash
Underfitting and overfitting
Q: What’s one way you can get underfitting?
Underfitting and overfitting

Q: What’s one way you can get underfitting?

A: A model that is too simple
Underfitting and overfitting

underfitting

overfitting
Q: What’s one way you can get overfitting?
Q: What’s one way you can get overfitting?

A: A model that is *too* complex (too many parameters)
Model, parameters and hyperparameters

Model: mathematical formulation of system (e.g., classifier)

Parameters: primary "knobs" of the model that are set by a learning algorithm

Hyperparameter: secondary "knobs"
A Terminology Buffet

the *task*: what kind of problem are you solving?

Classification
Regression
Clustering
A Terminology Buffet

Classification
- Regression
- Clustering

Fully-supervised
Semi-supervised
Un-supervised

*the task*: what kind of problem are you solving?

*the data*: amount of human input/number of labeled examples
A Terminology Buffet

- **Classification**
- **Regression**
- **Clustering**

- **Fully-supervised**
- **Semi-supervised**
- **Un-supervised**

**the task:** what kind of problem are you solving?

**the data:** amount of human input/number of labeled examples

**the approach:** how any data are being used

- Probabilistic
- Generative
- Conditional
- Spectral
- Neural
- Memory-based
- Exemplar
- ...
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today against a community in Junin department, central Peruvian mountain region.
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today against a community in Junin department, central Peruvian mountain region.
Electronic alerts have been used to assist the authorities in moments of chaos and potential danger: after the Boston bombing in 2013, when the Boston suspects were still at large, and last month in Los Angeles, during an active shooter scare at the airport.
Electronic alerts have been used to assist the authorities in moments of chaos and potential danger: after the Boston bombing in 2013, when the Boston suspects were still at large, and last month in Los Angeles, during an active shooter scare at the airport.
Classify with Goodness

\[ \text{best label} = \underset{\text{label}}{\text{arg max}} \text{ score(example, label)} \]
Classify with (Low) Regret/Loss

\[
\text{best label} = \arg \min_{\text{label}} \text{loss}(\text{example, label})
\]
Electronic alerts have been used to assist the authorities in moments of chaos and potential danger: after the Boston bombing in 2013, when the Boston suspects were still at large, and last month in Los Angeles, during an active shooter scare at the airport.
Classification Examples

Assigning subject categories, topics, or genres
Spam detection
Authorship identification

Age/gender identification
Language Identification
Sentiment analysis
...

...
Classification Examples

Assigning subject categories, topics, or genres
Spam detection
Authorship identification

Age/gender identification
Language Identification
Sentiment analysis
...

Input:

an instance
a fixed set of classes \( C = \{c_1, c_2, ..., c_J\} \)

Output: a predicted class \( c \) from \( C \)
Classification: Hand-coded Rules?

Assigning subject categories, topics, or genres
Spam detection
Authorship identification

Age/gender identification
Language Identification
Sentiment analysis

Rules based on combinations of words or other features
spam: black-list-address OR (“dollars” AND “have been selected”)

Accuracy can be high
If rules carefully refined by expert

Building and maintaining these rules is expensive

Can humans faithfully assign uncertainty?
Classification: Supervised Machine Learning

Assigning subject categories, topics, or genres
Spam detection
Authorship identification
Age/gender identification
Language Identification
Sentiment analysis

Input:
an instance $d$
a fixed set of classes $C = \{c_1, c_2, ..., c_J\}$
A training set of $m$ hand-labeled instances $(d_1, c_1), ..., (d_m, c_m)$

Output:
a learned classifier $\gamma$ that maps instances to classes
Classification: Supervised Machine Learning

Assigning subject categories, topics, or genres
Spam detection
Authorship identification

Age/gender identification
Language Identification
Sentiment analysis

... 

Input:
- an instance \( d \)
- a fixed set of classes \( C = \{c_1, c_2, \ldots, c_J\} \)
- A training set of \( m \) hand-labeled instances \( (d_1, c_1), \ldots, (d_m, c_m) \)

Output:
- a learned classifier \( \gamma \) that maps instances to classes

\( \gamma \) learns to associate certain features of instances with their labels
Classification: Supervised Machine Learning

Assigning subject categories, topics, or genres
Spam detection
Authorship identification
Age/gender identification
Language Identification
Sentiment analysis
...

**Input:**
- an instance $d$
- a fixed set of classes $C = \{c_1, c_2, ..., c_J\}$
- A training set of $m$ hand-labeled instances $(d_1, c_1), ..., (d_m, c_m)$

**Output:**
- a learned classifier $\gamma$ that maps instances to classes

Naïve Bayes
Logistic regression
Support-vector machines
k-Nearest Neighbors
...