ECE/CS/ME 532:
Matrix Methods in Machine Learning
(formerly Theory and Applications of Pattern Recognition)

Prof. Rebecca Willett (willett@discovery.wisc.edu)
NOT Rebekah Willett (willett@wisc.edu)
Machine Learning is about learning by example.

Ex. Imaging we want predict whether someone will enjoy the orig. Star Wars Trilogy.

First we take a survey of people who watched it; and ask

1. did you like it?
2. how much do you usually like Sci fi?
3. how much do you care about the Bechdel test?

Bechdel test: movie must have

- ≥ 2 Female characters (with names)
- talk to each other
- about something besides men
Then we plot the results

Can we now predict whether someone will like _Star Wars_ based on how much they like _Sci fi_ and the _Bechdel_ test?

One approach: predict label $y$ using a weighted combination of features $x_1 = \text{Sci fi pref.}$, $x_2 = \text{Bechdel pref.}$

$$y = \text{weight}_1 \cdot x_1 + \text{weight}_2 \cdot x_2$$
How do we choose weight₁ and weight₂ based on the survey results?

Let's index the surveys \( i = 1, 2, \ldots, n \) for \( i \)th survey, have label \( Y_i \) and features \( X_{i1} \) (sci fi) and \( X_{i2} \) (Bechdel).

Goal: make sure \( \hat{Y}_i = \text{weight}_1 X_{i1} + \text{weight}_2 X_{i2} \approx Y_i \) for all \( i \).
Matrix Methods help us find good weights.

Let \( \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \) be a vector of labels

Let \( \mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix} \)

\( x_{i,j} \) = survey response of \( i \)th person on \( j \)th question

each row of \( \mathbf{X} \) is response of one person

each column of \( \mathbf{X} \) is all responses to one question
vector of weights \( \mathbf{w} = \begin{bmatrix} \text{weight}_1 \\
\text{weight}_2 \\
\vdots \\
\text{weight}_p \end{bmatrix} \)

before \( \hat{y}_j = \sum_{j=1}^{p} \text{weight}_j \times j \)

now: \( \hat{y}_j = \mathbf{x} \cdot \mathbf{w} \Rightarrow \text{now we use matrix methods to find } \mathbf{w} \text{ so that } y_j \approx \hat{y}_j \)
Matrix methods are at the heart of

- machine learning
  - predicting labels/classification
  - recommender system
  - image recognition/analysis
- robotics
- finance
- mech/aero engineering
- signal processing
- optimization/operations research.

Beyond toolboxes:
- put algorithms in context/understand why they work
- design new algorithms
Machine Learning + Signal Processing Class @ Wisconsin

CS 540
intro to AI

CS 731
adv. AI

CS 760
machine learning

CS 539
ANN/fuzzy systems

CS/ECE 532
Matrix Methods ML

CS/ECE 7601
Math Foundations ML

CS/ECE 8601
Theoretical Foundations ML

ECE 901
Large-scale ML

ECE 730
prob. + stochastic processes

ECE 830
est. + detection theory

more applied survey style courses (lots of coding)

more mathematical focus (less coding)

Signal processing courses