CSci5521: Machine Learning Fundamentals

- Introduction

Slides from Prof. Catherine Zhao

Instructor

- Yoga Varatharajah

- Assistant Professor, Computer Science and Engineering

- Research Area: Machine Learning for Healthcare

- Contact: yvaratha@umn.edu

Teaching Assistants

- Haoyi Shi (shi00317@umn.edu)
- Xianhao Liu (liu03008@umn.edu)
- Drew Gjerstad (gjers043@umn.edu)
- Ryan Peters (pet03140@umn.edu)

Outline

- Course Overview
- Introduction to Machine Learning

Important Links

- Canvas Page: https://canvas.umn.edu/courses/541336

- Class website:
- https://yoga-varatha.github.io/csci5521-umn/

- Piazza: https://piazza.com/class/mklo5dw9uap1so

- Prerequisites:
  - Python programming
  - Statistics/probabilities, linear algebra, and some knowledge about multi-var calculus

- Lecture:
  - Tuesdays/Thursdays 09:45 AM - 11:00 AM
  - Vincent Hall 16

- Office hours:
  - Yoga: Thu 4pm-5pm, Keller 4-131
  - Haoyi: Mon 11am-12pm, Shepherd 234
  - Xianhao: Fri 3pm-4pm, Shepherd 344
  - Drew: Wed 2.30pm-3.30pm, Keller 1-211
  - Ryan: Fri 1.00pm-2.00pm, Keller 4-131

- Communication
  - All announcements will be made via canvas
  - All homework assignments will be managed through canvas
  - Use Piazza for all class-related questions/discussions

Grading Distribution

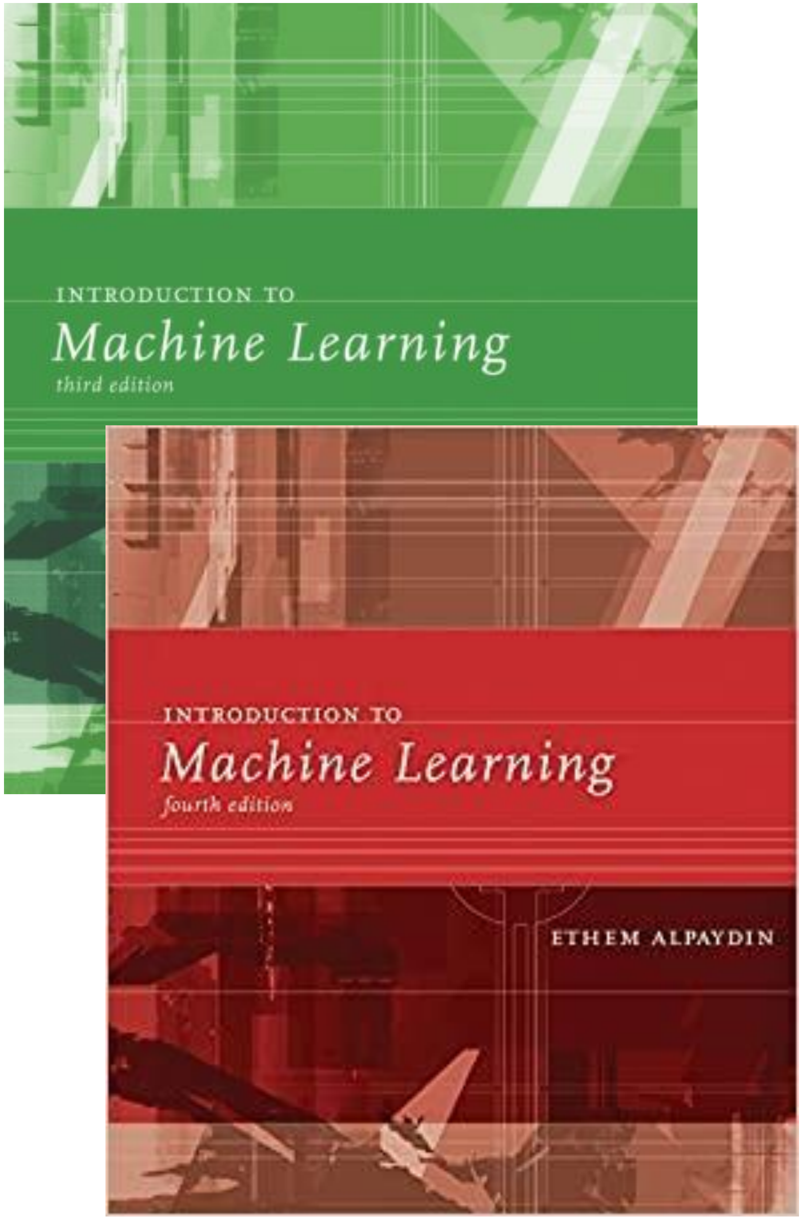
- Quizzes (~9%)
- Homework assignments (~56%)
- Midterm exam (~15%)
- Final exam (~20%)

- Quizzes
  - 4 quizzes
    - Best 3 out of 4 scores will be used
  - In class
  - Open book, with access to only materials from the class

- Homework Assignments
  - 5 homework assignments
    - First one will not count towards your grade but evaluate your background and readiness for the course.
  - All submissions are through canvas, due two weeks after posted (see schedule for tentative dates).
  - All programming in Python using specific libraries
  - Every late day will receive 10% penalty

- Midterm
  – In-class midterm
  – Tentative date: 3/24/26

- Final Exam
  – In-person
  – Date: 5/13/26



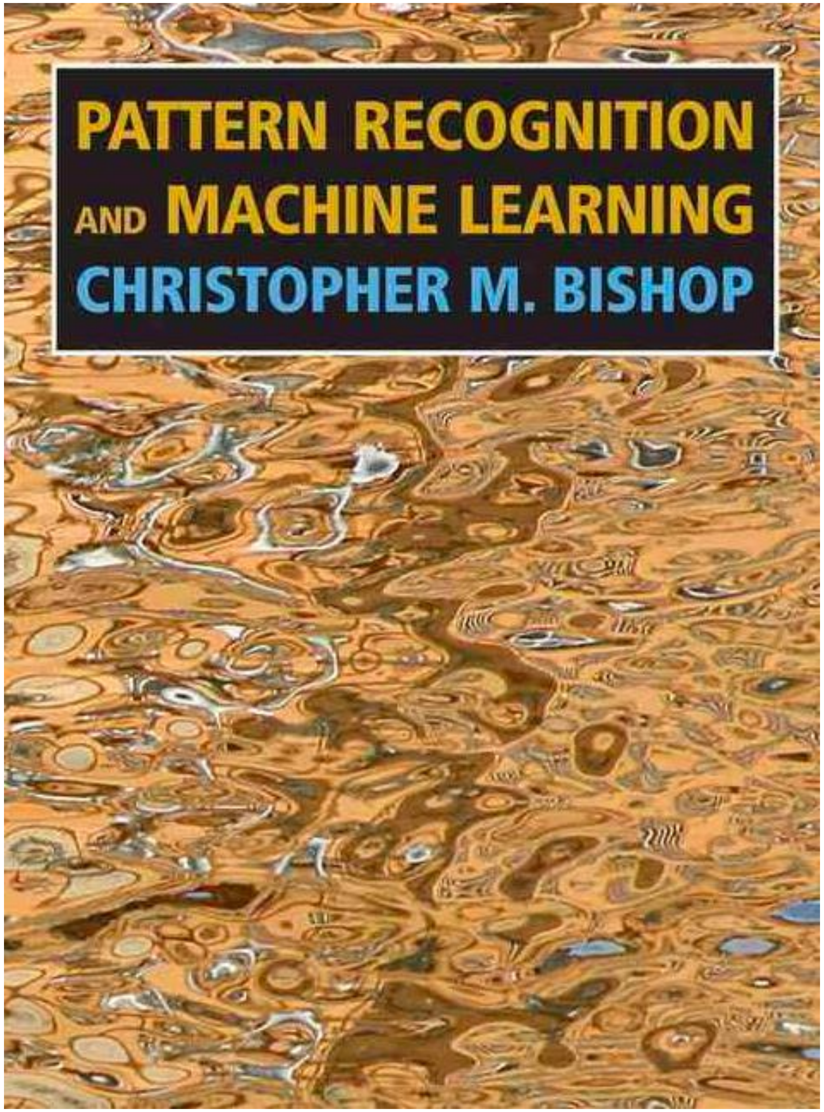
Introduction to Machine Learning,

Third / Fourth Edition

By Ethem Alpaydin

eBook:

http://login.ezproxy.lib.umn.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=nlebk&AN=2957329&site=ehost-live



Pattern Recognition and Machine Learning

By Christopher M. Bishop

https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

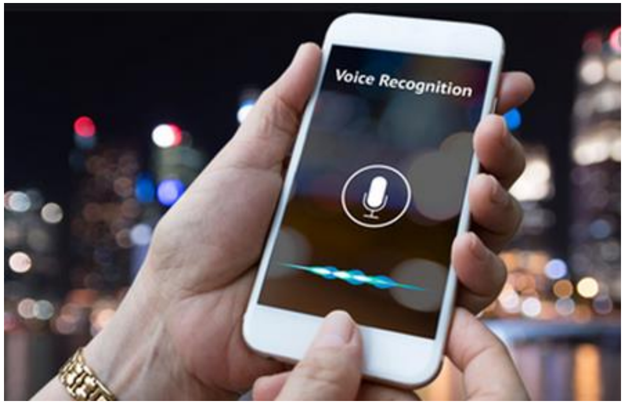
Policies

- Late homework submissions
  - Every late day will receive 10% penalty
  - Special circumstances require prior approval.

- Use of generative AI in assignments is NOT allowed.
  - If spotted, you will receive zero points on the assignment.

- Ongoing issues and safety concerns
  - Lectures of first week will be recorded and posted on canvas.
  - Starting second week, there will be a significant amount of chalk boarding which cannot be recorded. But lecture notes will be uploaded.

Machine Learning is Everywhere



Google a step closer to developing machines with human-like intelligence

Algorithms developed by Google designed to encode thoughts, could lead to computers with 'common sense' within a decade, says leading AI scientist

theguardian

ence More

ws / opinion / sport / arts / life

≡ Forbes

Mar 27, 2020, 02:47pm EDT | 6,874 views

Deep Learning: What You Need To Know

York Times

At iMerit offices in Kolkata, India, employees label images that are used to teach artificial intelligence systems. Rebecca Conway for The New York Times

A.I. Is Learning From Humans. Many Humans.

Artificial intelligence is being taught by thousands of office workers around the world. It is not exactly futuristic work.

nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 518 > Issue 7540 > News > Article

NATURE | NEWS

عربي

Game-playing software holds lessons for neuroscience

DeepMind computer provides new way to investigate how the brain works.

Elizabeth Gibney

'Deep learning' technology inspired by human brain

Machines are becoming increasingly more intelligent - able to see, speak and even think like us because of "deep learning", a set of algorithms that allows machines to see objects and understand what they are.

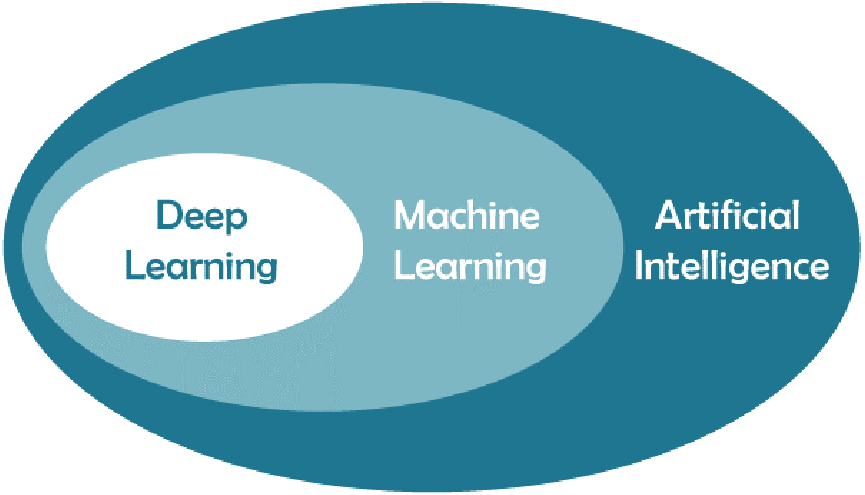
Yes, androids do dream of electric sheep

Google sets up feedback loop in its image recognition neural network - which looks for patterns in pictures - creating hallucinatory images of animals, buildings and landscapes which veer from beautiful to terrifying

- Retail
- Finance
- Manufacturing
- Medicine
- Bioinformatics
- Text understanding
- Speech recognition
- Computer vision
- Smart robot
- ……..

Artificial Intelligence (AI) & Machine Learning (ML)

- AI refers to the ability of computers to mimic human intelligence.
- ML is a way to achieve AI through learning from examples.

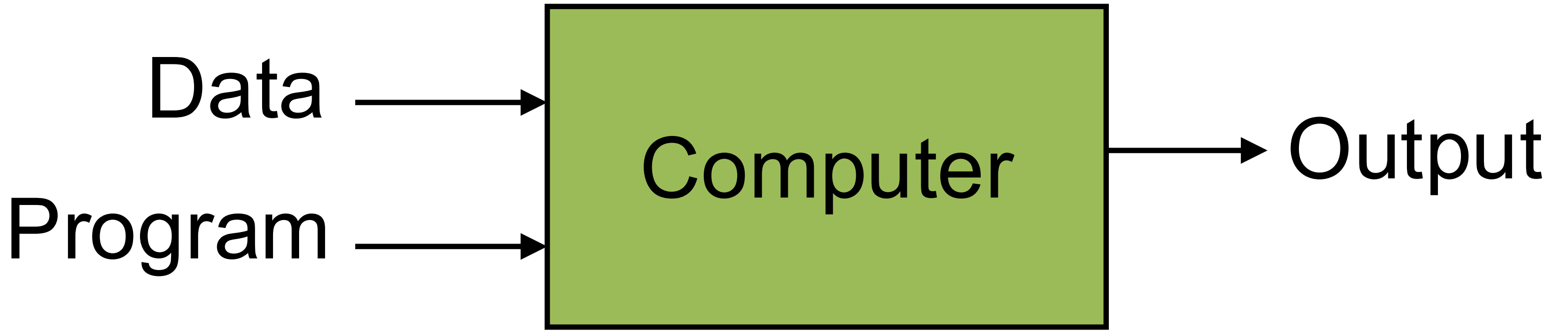


What is Machine Learning?

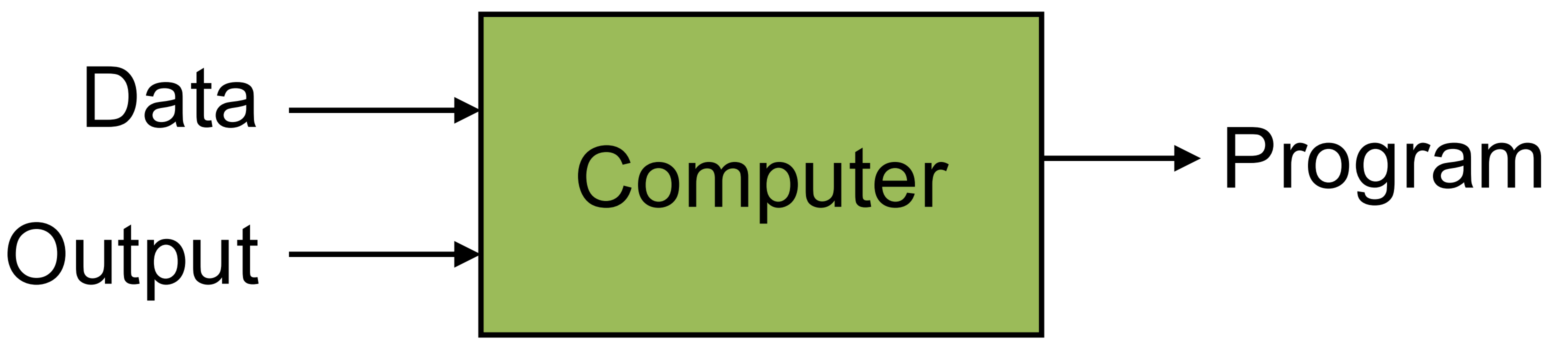
- Human learning vs machine learning
  - Humans learn from experience
  - Machines follow instructions

- How to make machines learn from examples? -- Machine Learning

Traditional Programming



Machine Learning



What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed."
-- Authur Samuel (1959)

"Machine Learning is the study of algorithms that
- improve their performance P
- at some task T
- with experience E.

A well defined learning task is given by <P, T, E>"
-- Tom Mitchell (1998)

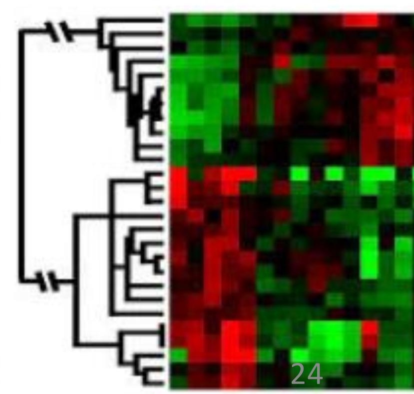
What is Machine Learning?

"Machine learning is programming computers to optimize a performance criterion using example data or past experience." -- Ethem Alpaydin

Why Machine Learning?

- Human expertise does not exist (navigating on Mars)
- Humans are unable to explain their expertise (speech recognition)
- Models need to be adapted to particular cases (personalized medicine)
- Models are based on huge amounts of data (genomics)

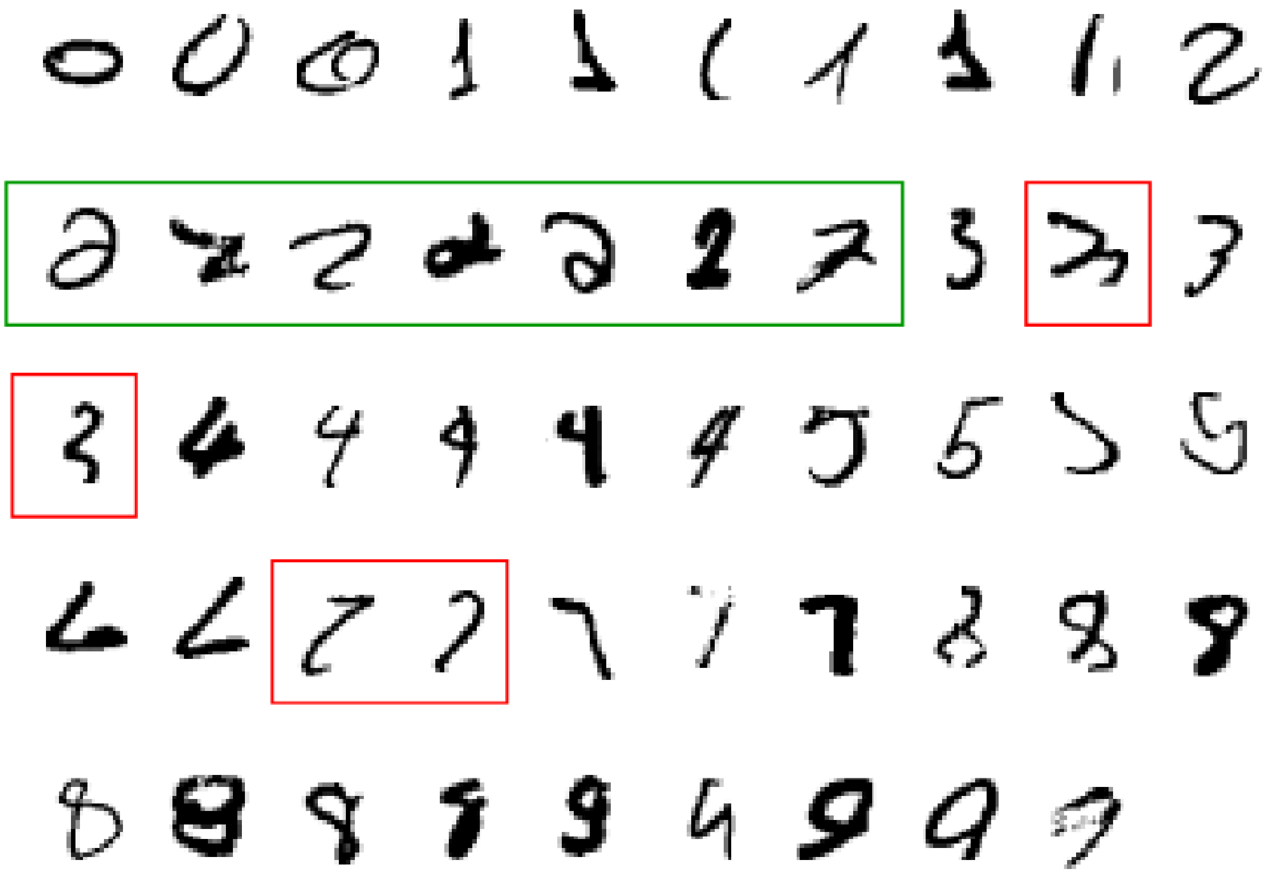


Learning is Not Always Useful

- Calculating square root?
- Linear regression?
- Calculating payroll?
- Recognizing digits?

A classic example of a task that requires machine learning. It is hard to say what makes a 2 or 3, etc.



Key Concepts

- Example data / training data: Observed instances possibly with known outcome

- Training / learning: Process of generalization from the example data as probability distributions, rules, discriminative functions...

- Performance criterion: Optimization of an objective function that evaluates the performance of solving a machine learning task

- Machine learning is data-driven
- Machine learning is one of the fastest growing fields in computer science
- Machine learning is important in both its own development and application domains

Notations

x	Scalar value
$\boldsymbol{x}$	Vector
$\boldsymbol{X}$	Matrix
$\mathbf{X}^T$	Transpose
$\mathbf{X}^{-1}$	Inverse
X	Random variable
$P(X)$	Probability mass function when X is discrete
$p(X)$	Probability density function when X is continuous
$P(X \backslash Y)$	Conditional probability of X given Y
$E[X]$	Expected value of the random variable X
Var (X)	Variance of X
Cov(X, Y)	Covariance of X and Y
Corr(X, Y)	Correlation of X and Y
μ	Mean
σ^2	Variance
Σ	Covariance matrix
m	Estimator to the mean
s^2	Estimator to the variance
S	Estimator to the covariance matrix
$\mathcal{N}(\mu, \sigma^2)$	Univariate normal distribution with mean n and variance σ2

Notations

$\mathcal{Z}$	Unit normal distribution: $\mathcal{N}$ (0,1)
$\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$	*d*-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
x	Input
d	Number of inputs (input dimensionality)
y	Output
r	Required output
K	Number of outputs (classes)
N	Number of training instances
z	Hidden value, intrinsic dimension, latent factor
k	Number of hidden dimensions, latent factors
C_i	Class i
X	Training sample
$\{x^t\}_{t=1}^N$	Set of x with index t ranging from 1 to N
$\{x^t, r^t\}_t$	Set of ordered pairs of input and desired output with index t
$g(x\|\theta)$	Function of x defined up to a set of parameters θ
argmax$\theta g(x\| \theta)$	The argument θ for which g has its maximum value
arg minθg $(x\|\theta)$	The argument θ for which g has its minimum value
$E(\theta\|X)$	Error function with parameters θ on the sample X
$l(\theta\|X)$	Likelihood of parameters θ on the sample X
$\mathcal{L}(\theta\|X)$	Log likelihood of parameters θ on the sample X
$1(c)$	1 if c is true, 0 otherwise
$\#\{c\}$	Number of elements for which c is true
δ_{ij}	Kronecker delta: 1 if $i = j$, 0 otherwise

- We will use a lot of notation
- Notation will be all over the place!
  - Every book, report, paper, etc., uses different notation
- Will use notation used in the text book
- Will use notation used in the (external) material we will use
- Clearly define notation before usage (ask if its unclear)

Topics

- Introduction
- Supervised learning
- Bayesian decision theory
- Parametric models
- Dimension reduction
- Clustering
- Nonparametric methods
- Linear discrimination
- Neural networks, deep learning
- Kernel machines
- Decision trees and random forests
- Graphical models
- Additional topics, time permitting..

Homework 0

- Will be released today or tomorrow at the latest!
- Required, but will not count towards grade
- Evaluates your readiness for this class
  - If you find it difficult, you will have to put in a lot of time to do well in this class
- Due in a week from when it is released

Questions?

- https://z.umn.edu/5521_exitpass

Some materials credit to Introduction to Machine Learning, by Ethem Alpaydin, and online resources