CS 1675: Intro to Machine Learning

Introduction

Prof. Adriana Kovashka
University of Pittsburgh
August 28, 2018
Course Info

• **Course website:**
  http://people.cs.pitt.edu/~kovashka/cs1675_fa18

• **Instructor:** Adriana Kovashka
  (kovashka@cs.pitt.edu)
  ➔ Use "CS1675" at the beginning of your Subject

• **Office:** Sennott Square 5325

• **Class:** Tue/Thu, 11am-12:15pm

• **Office hours:** Tue 2-3:55pm, Thu 1-3:55pm
About the Instructor

Born 1985 in Sofia, Bulgaria

Got BA in 2008 at Pomona College, CA (Computer Science & Media Studies)

Got PhD in 2014 at University of Texas at Austin (Computer Vision)
About the TA

- Karin Cox
- **Office**: Sennott Square 6150
- **Office hours**: TBD
  - Do the Doodle by the end of Friday: [https://doodle.com/poll/r8kbcezcrzuzuh8a](https://doodle.com/poll/r8kbcezcrzuzuh8a)
Recitations

• **Time:** Friday, 9am and 1pm
• **Room:** Sennott Square 6110
• **Instructor:** TBD
Course Goals

• To learn the basic machine learning techniques, both from a theoretical and practical perspective
• To practice implementing and using these techniques for simple problems
• To understand the advantages/disadvantages of machine learning algorithms and how they relate to each other
Textbooks


• More resources available on course webpage

• Your notes from class are your best study material, slides are not complete with notes
Programming Language

- We’ll use Matlab
- It can be downloaded for free from MyPitt
- We’ll do a short tutorial; ask TA if you need further help
Course Structure

• Lectures
• Weekly assignments
• Two exams
• Participation component
Policies and Schedule

http://people.cs.pitt.edu/~kovashka/cs1675_fa18
Should I take this class?

• It will be a lot of work!
  – I expect you’ll spend **6-8 hours** on HW each week
  – But you will learn a lot

• Some parts will be hard and require that you pay close attention!
  – But I will have periodic ungraded pop quizzes to see how you’re doing
  – I will also pick on students randomly to answer questions
  – Use instructor’s and TA’s office hours!!!
Questions?
Plan for Today

• Blitz introductions
• What is machine learning?
  – Example problems and tasks
  – ML in a nutshell
  – Challenges
  – Measuring performance
• Review
  – Linear algebra
  – Calculus
• Matlab tutorial
Blitz introductions (10 sec)

• What is your name?
• What one thing outside of school are you passionate about?
• What do you hope to get out of this class?

• Every time you speak, please remind me your name
What is machine learning?

• Finding patterns and relationships in data
• Using these patterns to make useful *predictions* or to *summarize* the data automatically

• E.g.
  – predict how much a user will like a movie, even though that user never rated that movie
  – identify common types of movies without knowing about genres
Example machine learning tasks

• Netflix challenge
  – Given lots of data about how users rated movies (training data)
  – But we don’t know how user $i$ will rate movie $j$ and want to predict that (test data)
  – Why is that hard? How can we do it?
Example machine learning tasks

- Spam or not?

CUSP has approved ECE 4424 with the following changes: please address all the issues addressed? (s sliding)

Thanks!!

Tracy

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going into business relationship with you. I am Nadia BamBa the only Daughter of late Mr and Mrs James BamBa, My father was a director of cocoa merchant in Abidjan, the economic capital of Ivory Coast before he was poisoned to death by his business associates on one of their outing to discuss on a business deal. When my mother died on the 21st October 2002, my father took me very special because i am motherless.

Before the death of my father in a private hospital here in Abidjan, He secretly called me on his bed side and told me that he had a sum of $6,800.000(SIX Million EIGHT HUNDRED THOUSAND), Dollars) left in a suspense account in a Bank here in Abidjan, that he used my name as his first Daughter for the next of kin in deposit of the fund.

He also explained to me that it was because of this wealth and some huge amount of money That his business associates supposed to balance him from the deal they had that he was poisoned by his business associates, that I should seek for a God fearing foreign partner in a country of my choice where I will transfer this money and use it for investment purposes, (such as real estate Or Hotel management),please i am honourably seeking your assistance in the following ways.

1) To provide a Bank account where this money would be transferred to.
2) To serve as the guardian of this Money since I am a girl of 19 years old.
3)Your private phone number’s and your family background’s that we can know each order more.

Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas.

Anticipating to hear from you soon.

Thanks and God Bless.

Best regards.
Example machine learning tasks

- Weather prediction

Slide credit: Carlos Guestrin
Example machine learning tasks

• Who will win <contest of your choice>?
Example machine learning tasks

- Machine translation

\[ x = \text{bringen sie bitte das auto zurück} \]

\[ y = \text{please return the car} \]
Example machine learning tasks

• Speech recognition
Example machine learning tasks

• Pose estimation
Example machine learning tasks

• Face recognition
Example machine learning tasks

• Image categorization

Slide credit: Dhruv Batra
Example machine learning tasks

• Image retrieval

Query: “black shoes”

Feedback: “more formal than these”

Feedback: “shinier than these”

Initial top search results

Refined top search results

Example machine learning tasks

• Inferring visual persuasion

Example machine learning tasks

- Answering questions about images

Example machine learning tasks

• What else?
• What are some problems from everyday life that can be helped by machine learning?
ML in a Nutshell

• Tens of thousands of machine learning algorithms

• Decades of ML research oversimplified:
  – Learn a mapping from input to output $f: X \rightarrow Y$
  – $X$: emails, $Y$: \{spam, notspam\}
ML in a Nutshell

\[ y = f(x) \]

- **Training**: given a *training set* of labeled examples \( \{(x_1,y_1), \ldots, (x_N,y_N)\} \), estimate the prediction function \( f \) by minimizing the prediction error on the training set.

- **Testing**: apply \( f \) to a never before seen *test example* \( x \) and output the predicted value \( y = f(x) \).
ML in a Nutshell

• Apply a prediction function to a feature representation of the image to get the desired output:

\[
\begin{align*}
f(\text{apple}) &= \text{“apple”} \\
f(\text{tomato}) &= \text{“tomato”} \\
f(\text{cow}) &= \text{“cow”}
\end{align*}
\]
ML in a Nutshell

Training Images

Training

Features

Training Labels

Learned model

Testing

Features

Learned model

Prediction

Test Image

Slide credit: D. Hoiem and L. Lazebnik
Training vs Testing

• What do we want?
  – High accuracy on training data?
  – No, high accuracy on *unseen/new/test data*!
  – Why is this tricky?

• Training data
  – Features (x) and labels (y) used to learn mapping f

• Test data
  – Features used to make a prediction
  – Labels only used to see how well we’ve learned f!!!

• Validation data
  – Held-out set of the *training data*
  – Can use both features and labels to tune *parameters* of the model we’re learning
Why do we hope this would work?

• Statistical estimation view:
  – x and y are *random variables*
  – \( D = (x_1,y_1), (x_2,y_2), \ldots, (x_N,y_N) \sim P(X,Y) \)
  – Both training & testing data sampled IID from \( P(X,Y) \)
  • IID: Independent and Identically Distributed
  – Learn on training set, have some hope of *generalizing* to test set

Adapted from Dhruv Batra
ML in a Nutshell

• Every machine learning algorithm has:
  – Data representation \((x, y)\)
  – Problem representation
  – Evaluation / objective function
  – Optimization

Adapted from Pedro Domingos
Data representation

• Let’s brainstorm what our “X” should be for various “Y” prediction tasks...
Problem representation

• Decision trees
• Sets of rules / Logic programs
• Instances
• Graphical models (Bayes/Markov nets)
• Neural networks
• Support vector machines
• Model ensembles
• Etc.
Evaluation / objective function

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Slide credit: Pedro Domingos
Optimization

• Discrete / combinatorial optimization
  – E.g. graph algorithms

• Continuous optimization
  – E.g. linear programming

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad Ax \leq b \\
\text{and} & \quad x \geq 0
\end{align*}
\]
Defining the Learning Task

Improve on task, \( T \), with respect to performance metric, \( P \), based on experience, \( E \).

\( T \): Categorize email messages as spam or legitimate
\( P \): Percentage of email messages correctly classified
\( E \): Database of emails, some with human-given labels

\( T \): Recognizing hand-written words
\( P \): Percentage of words correctly classified
\( E \): Database of human-labeled images of handwritten words

\( T \): Playing checkers
\( P \): Percentage of games won against an arbitrary opponent
\( E \): Playing practice games against itself

\( T \): Driving on four-lane highways using vision sensors
\( P \): Average distance traveled before a human-judged error
\( E \): A sequence of images and steering commands recorded while observing a human driver.
Types of Learning

• Supervised learning
  – Training data includes desired outputs

• Unsupervised learning
  – Training data does not include desired outputs

• Weakly or Semi-supervised learning
  – Training data includes a few desired outputs

• Reinforcement learning
  – Rewards from sequence of actions

Slide credit: Dhruv Batra
Types of Prediction Tasks

Supervised Learning

- **Classification**
  - $x$ \rightarrow \text{Classification} \rightarrow y\quad \text{Discrete}

- **Regression**
  - $x$ \rightarrow \text{Regression} \rightarrow y\quad \text{Continuous}

Unsupervised Learning

- **Clustering**
  - $x$ \rightarrow \text{Clustering} \rightarrow x'\quad \text{Discrete ID}

- **Dimensionality Reduction**
  - $x$ \rightarrow \text{Dimensionality Reduction} \rightarrow x'\quad \text{Continuous}

Adapted from Dhruv Batra
Navigating ML World

Machine Learning Algorithms Cheat Sheet

Unsupervised Learning: Clustering
- k-means
- k-modes
- Prefer Probability
- Categorical Variables
- Need to Specify k
- Hierarchical

Unsupervised Learning: Dimension Reduction
- Dimension Reduction
- Topic Modeling
- Probabilistic
- Latent Dirichlet Analysis

START

Supervised Learning: Classification
- Data Is Too Large
- Explainable
- Speed or Accuracy

Supervised Learning: Regression
- Speed or Accuracy

Decision Tree
- Random Forest
- Neural Network
- Gradient Boosting Tree

https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/
Example of Solving a ML Problem

• Spam or not?

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following changes: [Copy of the proposal with these items addressed]

Thanks!!!
Tracy

nadia bamba
To: undisclosed recipients;
Reply-To: nadia bamba
From Miss Nadia BamBa,

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going into a business relationship with you. I am Nadia BamBa the only Daughter of late Mr and Mrs James BamBa, My father was a director of cocoa merchant in Abidjan, the economic capital of Ivory Coast before he was poisoned to death by his business associates on one of their outing to discuss on a business deal. When my mother died on the 21st October 2002, my father took me very special because i am motherless.

Before the death of my father in a private hospital here in Abidjan, He secretly called me on his bedside and told me that he had a sum of $6,800,000(SIX Million EIGHT HUNDRED THOUSAND) Dollars left in a suspense account in a Bank here in Abidjan, that he used my name as his first Daughter for the next of kin in deposit of the fund.

He also explained to me that it was because of this wealth and some huge amount of money That his business associates supposed to balance him from the deal they had that he was poisoned by his business associates, that I should seek for a God fearing foreign partner in a country of my choice where I will transfer this money and use it for investment purposes, (such as real estate Or Hotel management).please i am honourably seeking your assistance in the following ways.

1) To provide a Bank account where this money would be transferred to.
2) To serve as the guardian of this Money since I am a girl of 19 years old.
3)Your private phone number’s and your family background’s that we can know each other more.

Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas.

Anticipating to hear from you soon.
Thanks and God Bless.
Best regards.

VS
Intuition

• Spam Emails
  – a lot of words like
    • “money”
    • “free”
    • “bank account”

• Regular Emails
  – word usage pattern is more spread out
Simple strategy: Let’s count!

This is X

\[
\begin{pmatrix}
\text{free} & 100 \\
\text{money} & 2 \\
\vdots & \vdots \\
\text{account} & 2 \\
\vdots & \vdots
\end{pmatrix}
\]

This is Y

\[
\text{SPAM}
\]

= 1 or 0?

Adapted from Dhruv Batra, Fei Sha

nadia bamba
To: undisclosed recipients;
Reply-To: nadia bamba
From Miss Nadia BamBa,

From Miss Nadia BamBa,

Greeting, Permit me to inform you of my desire of going i Nadia BamBa the only Daughter of late Mr and Mrs Jame cocoa merchant in Abidjan, the economic capital of Ivory his business associates on one of their outing to discuss on the 21st October 2002, my father took me very speci

Before the death of my father in a private hospital here ii bedside and told me that he had a sum of $6, 8000.000(S Dollars) left in a suspense account in a Bank here in Abic Daughter for the next of kin in deposit of the fund.

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following changes: Can copy of the proposal with these items addressed? (see below) Thanks!!! Tracy
Weigh counts and sum to get prediction

Adapted from Dhruv Batra, Fei Sha
Why not just hand-code these weights?

• We’re letting the data do the work rather than develop hand-code classification rules
  – The *machine* is *learning* to program itself

• But there are challenges...
Klingon vs Mlingon Classification

• Training Data
  – Klingon: klix, kour, koop
  – Mlingon: moo, maa, mou

• Testing Data: kap

• Which language? Why?

Slide credit: Dhruv Batra
“I saw her duck”
“I saw her duck”
“I saw her duck”
“I saw her duck with a telescope...”
What humans see
What computers see
Challenges

• Some challenges: ambiguity and context
• Machines take data representations too literally
• Humans are much better than machines at generalization, which is needed since test data will rarely look exactly like the training data
Challenges

• Why might it be hard to:
  – Predict if a viewer will like a movie?
  – Recognize cars in images?
  – Translate between languages?
The Time is Ripe to Study ML

• Many basic effective and efficient algorithms available.
• Large amounts of on-line data available.
• Large amounts of computational resources available.
Where does ML fit in?

- **Psychology Physiology**
  - biology of learning
  - inspiring paradigms
  - Ex: neural networks

- **Applied Maths**
  - optimization
  - linear algebra
  - Ex: convex optim

- **Applications**
  - new challenges
  - Ex: ad placement

- **Computer Science**
  - algorithm design
  - data structure
  - complexity analysis
  - Ex: kd tree

- **Statistics**
  - estimation techniques
  - theoretical framework
  - optimality, efficiency
  - Ex: learning theory

Slide credit: Dhruv Batra, Fei Sha
Plan for Today

• Blitz introductions
• What is machine learning?
  – Example problems and tasks
  – ML in a nutshell
  – Challenges
  – Measuring performance
• Review
  – Linear algebra
  – Calculus
• Matlab tutorial
Measuring Performance

• If $y$ is discrete:
  – Accuracy: $\frac{\# \text{ correctly classified}}{\# \text{ all test examples}}$
  – Precision/recall
    • True Positive, False Positive, True Negative, False Negative
    • $\text{Precision} = \frac{\# \text{ predicted true pos}}{\# \text{ predicted pos}}$ = $\frac{TP}{TP + FP}$
    • $\text{Recall} = \frac{\# \text{ predicted true pos}}{\# \text{ true pos}}$ = $\frac{TP}{TP + FN}$
  – F-measure
    = $\frac{2PR}{P + R}$

• Want evaluation metric to be in some range, e.g. [0 1]
  – 0 = worst possible classifier, 1 = best possible classifier
Precision / Recall / F-measure

**True positives**
(images *that contain* people)

- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F-measure

**True negatives**
(images *do not contain* people)

**Predicted positives**
(images *predicted to contain* people)

- Accuracy: 5 / 10 = 0.5

**Predicted negatives**
(images *predicted not to contain* people)
Measuring Performance

• If $y$ is continuous:
  – Sum-of-Squared-Differences (SSD) error between predicted and true $y$:

$$E = \sum_{i=1}^{n} \left( f(x_i) - y_i \right)^2$$
Linear algebra review

See http://cs229.stanford.edu/section/cs229-linalg.pdf for more
Vectors and Matrices

- Vectors and matrices are just collections of ordered numbers that represent something: movements in space, scaling factors, word counts, movie ratings, pixel brightnesses, etc.
- We’ll define some common uses and standard operations on them.
Vector

• A column vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ where

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

• A row vector $\mathbf{v}^T \in \mathbb{R}^{1 \times n}$ where

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$$

$T$ denotes the transpose operation
Vector

• You’ll want to keep track of the orientation of your vectors when programming in MATLAB.
• You can transpose a vector $V$ in MATLAB by writing $V'$. 
Vectors have two main uses

- Vectors can represent an offset in 2D or 3D space
- Points are just vectors from the origin
- Data can also be treated as a vector
- Such vectors don’t have a geometric interpretation, but calculations like “distance” still have value
Matrix

• A matrix \( \mathbf{A} \in \mathbb{R}^{m \times n} \) is an array of numbers with size \( m \downarrow \) by \( n \rightarrow \), i.e., \( m \) rows and \( n \) columns.

\[
\mathbf{A} = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn}
\end{bmatrix}
\]

• If \( m = n \), we say that \( \mathbf{A} \) is square.
Matrix Operations

• Addition

\[
\begin{bmatrix}
a & b \\
c & d \\
\end{bmatrix} + \begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix} = \begin{bmatrix}
a + 1 & b + 2 \\
c + 3 & d + 4 \\
\end{bmatrix}
\]

– Can only add a matrix with matching dimensions, or a scalar.

\[
\begin{bmatrix}
a & b \\
c & d \\
\end{bmatrix} + 7 = \begin{bmatrix}
a + 7 & b + 7 \\
c + 7 & d + 7 \\
\end{bmatrix}
\]

• Scaling

\[
\begin{bmatrix}
a & b \\
c & d \\
\end{bmatrix} \times 3 = \begin{bmatrix}
3a & 3b \\
3c & 3d \\
\end{bmatrix}
\]
Matrix Multiplication

• Let X be an $axb$ matrix, Y be an $bxc$ matrix
• Then Z = X*Y is an $axc$ matrix
• Second dimension of first matrix, and first dimension of first matrix have to be the same, for matrix multiplication to be possible
• Practice: Let X be an 10x5 matrix. Let’s factorize it into 3 matrices...
Matrix Multiplication

• The product $AB$ is:

• Each entry in the result is (that row of $A$) dot product with (that column of $B$)
Matrix Multiplication

• Example:

$$\begin{bmatrix} 0 & 2 \\ 4 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} \_ & \_ \\ \_ & \_ \end{bmatrix}$$

- Each entry of the matrix product is made by taking the dot product of the corresponding row in the left matrix, with the corresponding column in the right one.

$$0 \cdot 3 + 2 \cdot 7 = 14$$
Inner Product

• Multiply corresponding entries of two vectors and add up the result

\[ \mathbf{x}^T \mathbf{y} = [x_1 \ldots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \quad \text{(scalar)} \]

• \( \mathbf{x} \cdot \mathbf{y} \) is also \( |\mathbf{x}| \ |\mathbf{y}| \cos(\text{angle between } \mathbf{x} \text{ and } \mathbf{y}) \)

• If \( \mathbf{B} \) is a unit vector, then \( \mathbf{A} \cdot \mathbf{B} \) gives the length of \( \mathbf{A} \) which lies in the direction of \( \mathbf{B} \) (projection)

\( \theta \)
Different types of product

- $\mathbf{x}, \mathbf{y}$ = column vectors (nx1)
- $\mathbf{X}, \mathbf{Y}$ = matrices (mxn)
- $\mathbf{x}, \mathbf{y}$ = scalars (1x1)

- $\mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \text{inner product (1xn x nx1 = scalar)}$
- $\mathbf{x} \otimes \mathbf{y} = \mathbf{x} \mathbf{y}^T = \text{outer product (nx1 x 1xn = matrix)}$

- $\mathbf{X} \ast \mathbf{Y} = \text{matrix product}$
- $\mathbf{X} \ast \mathbf{Y} = \text{element-wise product}$
Inverse

• Given a matrix $A$, its inverse $A^{-1}$ is a matrix such that $AA^{-1} = A^{-1}A = I$

• E.g. $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$

• Inverse does not always exist. If $A^{-1}$ exists, $A$ is invertible or non-singular. Otherwise, it’s singular.
Matrix Operations

• Transpose – flip matrix, so row 1 becomes column 1

\[
\begin{bmatrix}
0 & 1 & \ldots
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 & 1 \\
2 & 3 \\
4 & 5
\end{bmatrix}^T = \begin{bmatrix}
0 & 2 & 4 \\
1 & 3 & 5
\end{bmatrix}
\]

• A useful identity:

\[
(ABC)^T = C^T B^T A^T
\]
Norms

• L1 norm

\[ \| x \|_1 := \sum_{i=1}^{n} |x_i| \]

• L2 norm

\[ \| x \| := \sqrt{x_1^2 + \cdots + x_n^2} \]

• L^p norm (for real numbers \( p \geq 1 \))

\[ \| x \|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \]
Matrix Rank

• Column/row rank
  \[ \text{col-rank}(A) = \text{the maximum number of linearly independent column vectors of } A \]
  \[ \text{row-rank}(A) = \text{the maximum number of linearly independent row vectors of } A \]

• Column rank always equals row rank
• Matrix rank \[ \text{rank}(A) \triangleq \text{col-rank}(A) = \text{row-rank}(A) \]
• If a matrix is not full rank, inverse doesn’t exist
  – Inverse also doesn’t exist for non-square matrices
Matrix Operation Properties

• Matrix addition is commutative and associative
  - $A + B = B + A$
  - $A + (B + C) = (A + B) + C$

• Matrix multiplication is associative and distributive but *not* commutative
  - $A(B* C) = (A*B)C$
  - $A(B + C) = A*B + A*C$
  - $A*B \neq B*A$
Special Matrices

- **Identity matrix $I$**
  - Square matrix, 1’s along diagonal, 0’s elsewhere
  - $I \cdot \text{[another matrix]} = \text{[that matrix]}$

- **Diagonal matrix**
  - Square matrix with numbers along diagonal, 0’s elsewhere
  - A diagonal $\cdot \text{[another matrix]}$ scales the rows of that matrix

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
3 & 0 & 0 \\
0 & 7 & 0 \\
0 & 0 & 2.5 \\
\end{bmatrix}
\]
Special Matrices

• Symmetric matrix

\[ A^T = A \]

\[
\begin{bmatrix}
1 & 2 & 5 \\
2 & 1 & 7 \\
5 & 7 & 1
\end{bmatrix}
\]
Matrix Operations

- MATLAB example:

\[ AX = B \]

\[ A = \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \]

\[
\begin{align*}
>> \quad & x = A \backslash B \\
& x = \\
& \begin{bmatrix} 1.0000 \\ -0.5000 \end{bmatrix}
\end{align*}
\]
Linear independence

• Suppose we have a set of vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_n \)
• If we can express \( \mathbf{v}_1 \) as a linear combination of the other vectors \( \mathbf{v}_2 \ldots \mathbf{v}_n \), then \( \mathbf{v}_1 \) is linearly dependent on the other vectors.
  – The direction \( \mathbf{v}_1 \) can be expressed as a combination of the directions \( \mathbf{v}_2 \ldots \mathbf{v}_n \). (E.g. \( \mathbf{v}_1 = 0.7 \mathbf{v}_2 - 0.7 \mathbf{v}_4 \))
• If no vector is linearly dependent on the rest of the set, the set is linearly independent.
  – Common case: a set of vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_n \) is always linearly independent if each vector is perpendicular to every other vector (and non-zero)
Linear independence

Linearly independent set  Not linearly independent
Singular Value Decomposition (SVD)

• There are several computer algorithms that can “factor” a matrix, representing it as the product of some other matrices.
• The most useful of these is the Singular Value Decomposition.
• Represents any matrix $A$ as a product of three matrices: $U \Sigma V^T$
• MATLAB command: $[U,S,V] = \text{svd}(A)$;
Singular Value Decomposition (SVD)

\[ U \Sigma V^T = A \]

- Where \( U \) and \( V \) are rotation matrices, and \( \Sigma \) is a scaling matrix. For example:

\[
\begin{bmatrix}
-0.40 & 0.916 \\
0.916 & 0.40
\end{bmatrix}
\times
\begin{bmatrix}
5.39 & 0 \\
0 & 3.154
\end{bmatrix}
\times
\begin{bmatrix}
-0.05 & 0.999 \\
0.999 & 0.05
\end{bmatrix}
= \begin{bmatrix}
3 & -2 \\
1 & 5
\end{bmatrix}
\]
Singular Value Decomposition (SVD)

- In general, if $A$ is $m \times n$, then $U$ will be $m \times m$, $\Sigma$ will be $m \times n$, and $V^T$ will be $n \times n$.

\[
\begin{bmatrix}
-0.39 & -0.92 \\
-0.92 & 0.39
\end{bmatrix}
\begin{bmatrix}
9.51 & 0 & 0 \\
0 & 0.77 & 0
\end{bmatrix}
\begin{bmatrix}
-0.42 & -0.57 & -0.70 \\
0.81 & 0.11 & -0.58 \\
0.41 & -0.82 & 0.41
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\]
Singular Value Decomposition (SVD)

• **U** and **V** are always rotation matrices.
  – Geometric rotation may not be an applicable concept, depending on the matrix. So we call them “unitary” matrices – each column is a unit vector.

• **Σ** is a diagonal matrix
  – The number of nonzero entries = rank of **A**
  – The algorithm always sorts the entries high to low

\[
U = \begin{bmatrix}
  -.39 & -.92 \\
  -.92 & .39 \\
\end{bmatrix} \times \begin{bmatrix}
  9.51 & 0 & 0 \\
  0 & .77 & 0 \\
\end{bmatrix} \times \begin{bmatrix}
  -.42 & -.57 & -.70 \\
  .81 & .11 & -.58 \\
  .41 & -.82 & .41 \\
\end{bmatrix} = \begin{bmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6 \\
\end{bmatrix}
\]
Singular Value Decomposition (SVD)

\[ M = U \Sigma V^T \]
Calculus review
Differentiation

The derivative provides us information about the rate of change of a function.

The derivative of a function is also a function.

Example:
The derivative of the rate function is the acceleration function.
Derivative = rate of change
Derivative = rate of change

- Linear function $y = mx + b$
- Slope $m = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x}$,
Ways to Write the Derivative

Given the function $f(x)$, we can write its derivative in the following ways:

- $f'(x)$

- $\frac{df}{dx}(x)$

The derivative of $x$ is commonly written $dx$. 

Texas A&M Dept of Statistics
Differentiation Formulas

The following are common differentiation formulas:

- The derivative of a constant is 0.
  \[
  \frac{d}{du} c = 0
  \]

- The derivative of a sum is the sum of the derivatives.
  \[
  \frac{d}{du} (f(u) + g(u)) = f'(u) + g'(u)
  \]
Examples

- The derivative of a constant is 0.

\[ \frac{d}{du} 7 = \]

- The derivative of a sum is the sum of the derivatives.

\[ \frac{d}{dt} (t + 4) = \]

Texas A&M Dept of Statistics
More Formulas

- The derivative of $u$ to a constant power:
  \[
  \frac{d}{du} u^n = n \cdot u^{n-1} \, du
  \]
- The derivative of $e$:
  \[
  \frac{d}{du} e^u = e^u \, du
  \]
- The derivative of $\log$:
  \[
  \frac{d}{du} \log(u) = \frac{1}{u} \, du
  \]
More Examples

- The derivative of \( u \) to a constant power:

\[
\frac{d}{dx} 3x^3 =
\]

- The derivative of \( e \):

\[
\frac{d}{dy} e^4y =
\]

- The derivative of \( \log \):

\[
\frac{d}{dx} 3\log(x) =
\]
Product and Quotient

The product rule and quotient rules are commonly used in differentiation.

- **Product rule:**
  \[
  \frac{d}{du} (f(u) \cdot g(u)) = f(u)g'(u) + g(u)f'(u)
  \]

- **Quotient rule:**
  \[
  \frac{d}{du} \left( \frac{f(u)}{g(u)} \right) = \frac{g(u)f'(u) - f(u)g'(u)}{(g(u))^2}
  \]
The chain rule allows you to combine any of the differentiation rules we have already covered.

- First, do the derivative of the outside and then do the derivative of the inside.

\[
\frac{d}{du} f(g(u)) = f'(g(u)) \times g'(u) \times du
\]
Try These

\[ f(z) = z + 11 \]
\[ s(y) = 4ye^{2y} \]

\[ g(y) = 4y^3 + 2y \]
\[ p(x) = \frac{\log(x^2)}{x} \]

\[ h(x) = e^{3x} \]
\[ q(z) = (e^z - z)^3 \]
Solutions

\[ f'(z) = 1 \]
\[ s'(y) = 8ye^{2y} + 4e^{2y} \]

\[ g'(y) = 12y^2 + 2 \]
\[ p'(x) = \frac{2 - \log(x^2)}{x^2} \]

\[ h'(x) = 3e^{3x} \]
\[ q'(z) = 3(e^z - z)^2(e^z - 1) \]
Matlab
Matlab tutorial

http://www.cs.pitt.edu/~kovashka/cs1675_fa18/tutorial.m
http://www.cs.pitt.edu/~kovashka/cs1675_fa18/myfunction.m
http://www.cs.pitt.edu/~kovashka/cs1675_fa18/myotherfunction.m

Please cover whatever we don’t finish at home.
Other tutorials and exercises

• [Link](https://people.cs.pitt.edu/~milos/courses/cs2750/Tutorial/)
• [Link](http://www.math.udel.edu/~braun/M349/Matlab_probs2.pdf)
• [Link](http://www.facstaff.bucknell.edu/maneval/help211/basicexercises.html)
  – Do Problems 1-8, 12
  – Most also have solutions
  – Ask the TA if you have any problems