CIT651 – Introduction to Machine Learning and Statistical Data Analysis

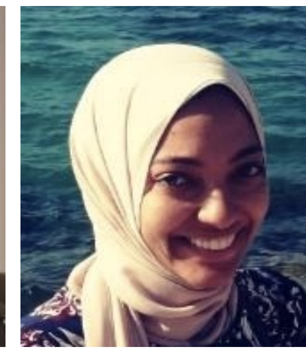
Mustafa Elattar

Spring 2024

Nile University

CIT651 – Introduction to Machine Learning and Statistical Data Analysis

- Instructor:
  Dr. Mustafa A. Elattar, Associate Professor
  melattar@nu.edu.eg
  Location: UB1 – 210



- Teacher Assistants:
  Eng. Shahenda Hatem
  shatem@nu.edu.eg

- References
  - Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer, 2006
  - Supporting materials
  - Research papers

Calendar and Syllabus (Initial)

Date	Topic	Lab	Assignments
4/3	Topic 1: Course Introduction, Linear Algebra Revision		
11/3	Topic 2: Probability Theory Revision	Yes	
18/3	Topic 3: Statistics Basics		Assignment 1,2 Release
25/3	Topic 4: Regression Models	Yes	Assignment 3 Release
1/4	Topic 5: Linear Classifiers 1 – Linear Classifiers	Yes	Assignment 1,2 Deadline
8/4	Topic 6: Text and Image Feature Extraction		Assignment 3 Deadline
15/4	Topic 7: Linear Classifiers 2 - Probabilistic Models	Yes	Project registered for each three students
22/4	Topic 8: Linear Classifiers 3 - Support Vector Machines	Yes	Assignment 4 Release
29/4	Topic 9: Nonlinear Classifiers		Assignment 4 Deadline Assignment 5 Release
6/5	Sham Nesim Holiday		
13/5	Topic 10: Data Clustering 1	Yes	Assignment 5 Deadline Assignment 6 Release
20/5	Topic 11: Data Clustering 2	Yes	Assignment 6 Deadline
27/5	Topic 12: Principal Components Analysis		
3/6	Project 2 Submission, Discussion, and Oral Presentation		
10/6	Final Examination		
17/6			
23/6	Grade Submission Deadline		

CIT651 – Introduction to Machine Learning and Statistical Data Analysis

Grading Policy: (this grade distribution is subject to change):

- Attendance (5%)
- Assignments (40%)
  - Stat. Assignment 1 5%
  - Stat. Assignment 2 5%
  - Stat. Assignment 3 5%
  - ML Assignment 1 8%
  - ML Assignment 2 8%
  - ML Assignment 3 9%
- Quiz 5%
- Project 1 5%
- Project 2 15%
- Final Exam 30%

Machine Learning?

Informal definition (Wikipedia):

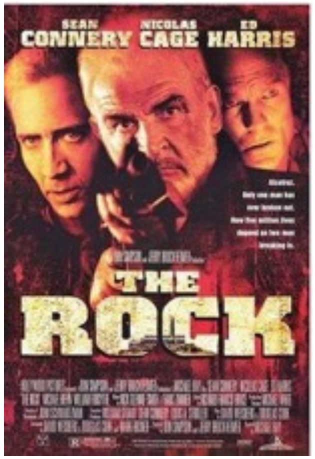
Designing and developing algorithms that allow computers to evolve behaviors based on empirical data

Machine Learning?

Formal definition:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E

Examples: Movie Recommendation System



Watch History (Experience E)

Recommendation

(Task T)

Examples:

- Netflix
- Youtube.com

Does it match the user's preference?

(Performance P)

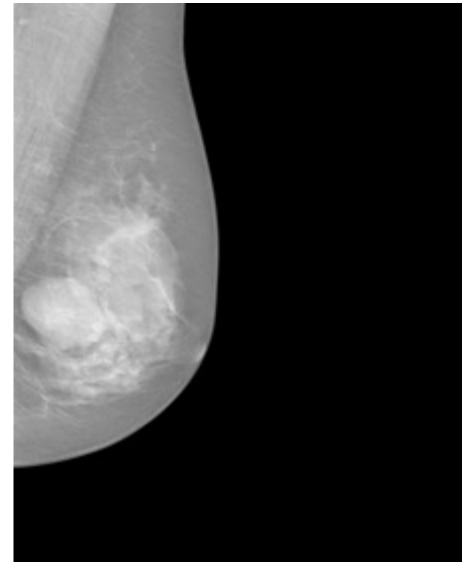
Examples: Cancer Identification

This image have normal and homogenous breast densities

Normal

This image features normal and homogenous breast densities

Normal



There are multiple dense areas located in the MLO view

Abnormal

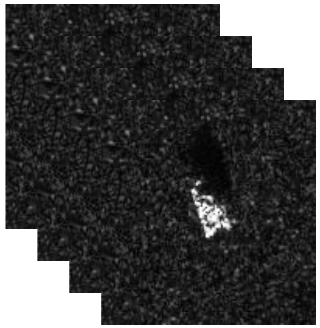
Experience E

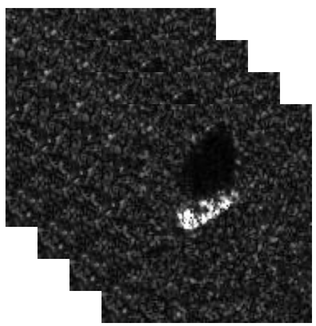
Task T

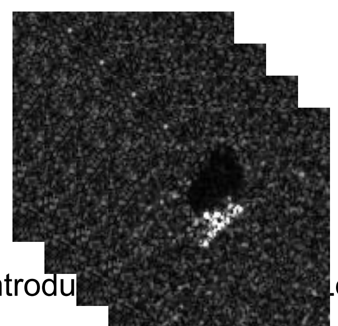
Performance P ?

Examples: Automatic Target Recognition (ATR)

Training Dataset
(Experience E)

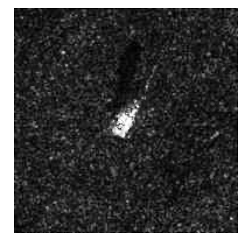







Which target is it?

(Task T)

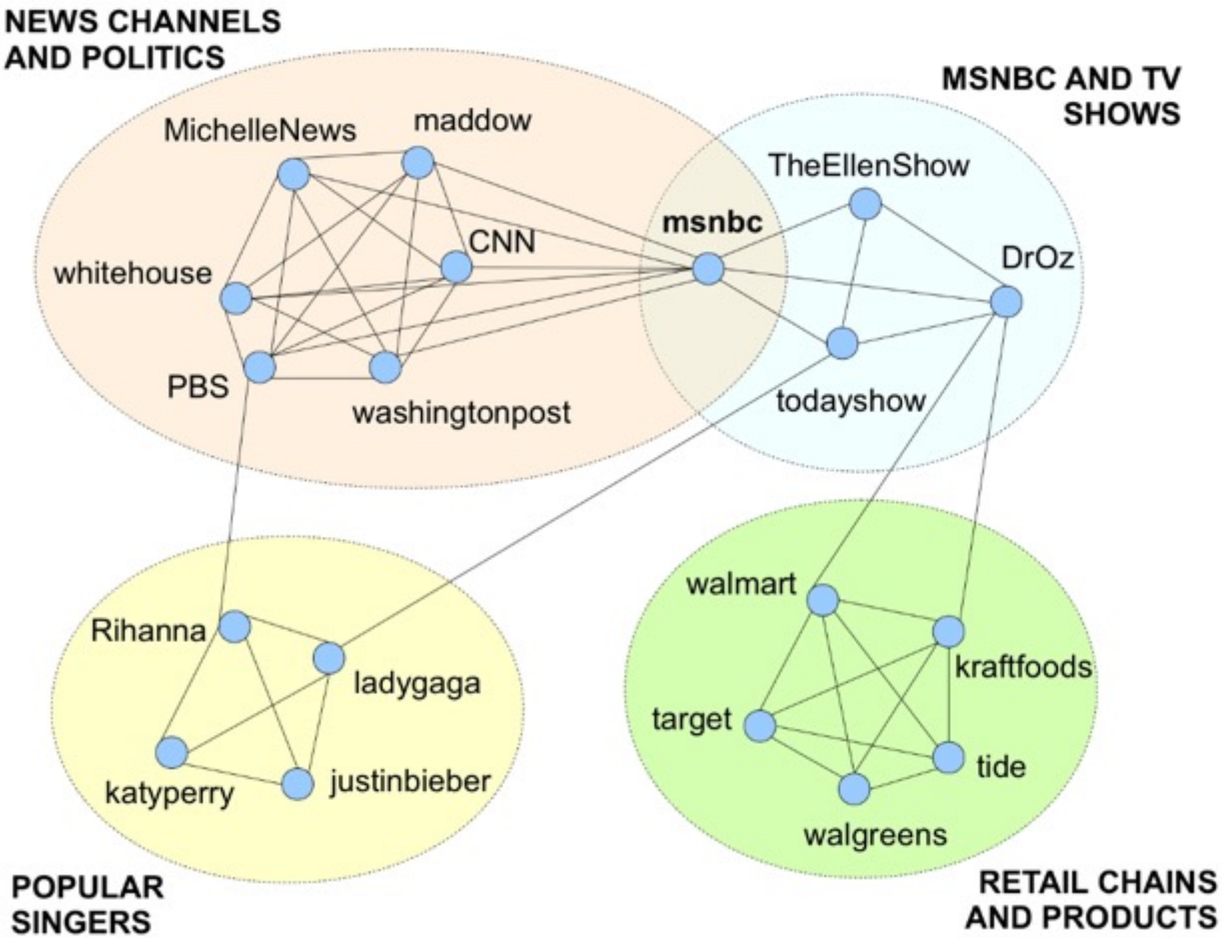


How many targets are classified correctly?

(Performance P)

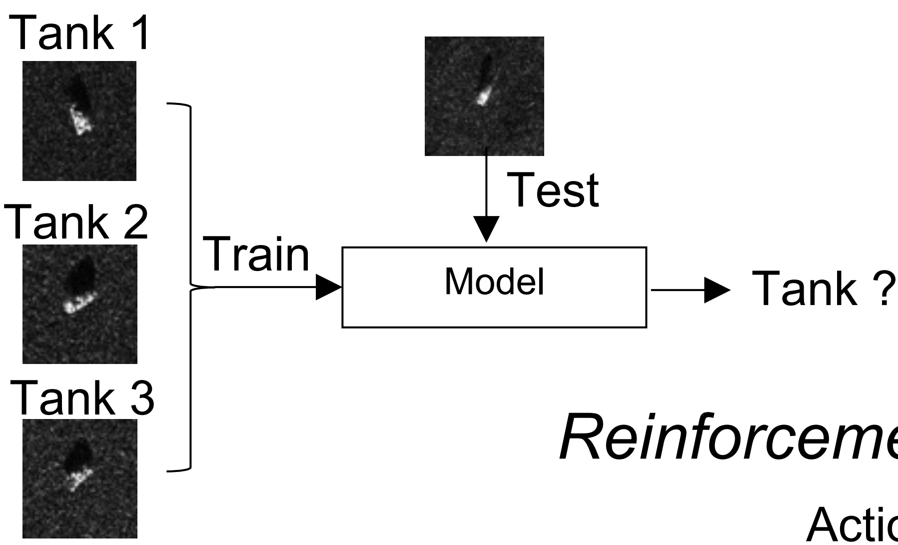
Examples: Social Network Analysis

- Grouping communities together

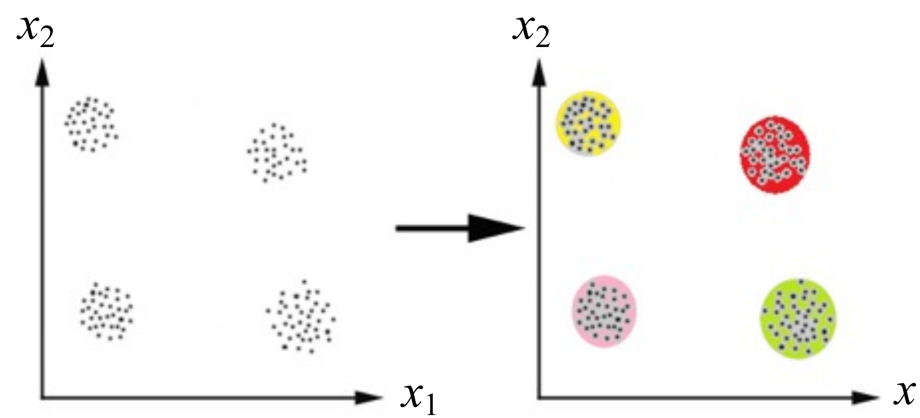


Machine Learning Algorithms

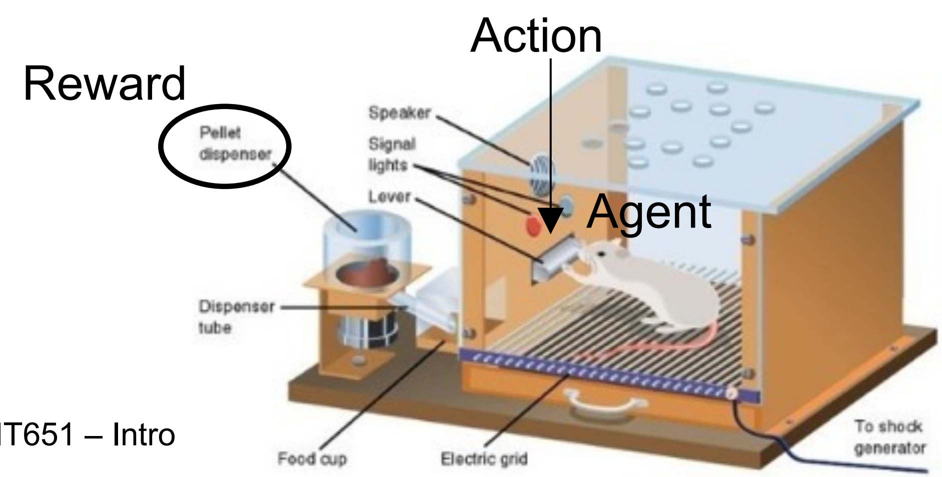
Supervised Learning



Unsupervised Learning



Reinforcement Learning



Course Outline

- Linear Algebra Review
- Probability Theory Review
- Statistical Parameter Estimation
- Hypothesis Testing
- Regression Analysis
  - Simple Linear Regression
  - Multiple Linear Regression
- Linear Classification:
  - Discriminant Functions:
    - Least Squares Classifier
    - Fisher's Linear Discriminant
    - Perceptron
    - Support Vector Machine
  - Probabilistic Generative Models:
    - Gaussian Generative Model
    - Naïve Bayes Classifier
  - Probabilistic Discriminative Models:
    - Logistic Regression
    - Evaluation Metrics
- **Text Feature Extraction**
- **Visual Feature Extraction**
- **Non-linear Classification:**
  - Instance-based Learning:
    - K-nearest Neighbor Classifier
    - Weighted K-nearest Neighbor
  - Decision Tree Learning
- **Clustering Techniques:**
  - K-means Clustering
  - Clustering Validity Indices
  - Fuzzy C-means Clustering
  - Hierarchical Clustering
  - DBSCAN Clustering
- **Dimensionality Reduction and Feature Extraction:**
  - Principal Component Analysis
  - Independent Component Analysis

Lec 1 - Linear Algebra Review

Mustafa Elattar

Nile University

Linear Algebra Review: Matrices

- Matrix: A set of elements organized in rows and columns

$$\begin{array}{cc} & \begin{array}{cc}\text{Col 1} & \text{Col 2}\end{array} \\ \begin{array}{c}\text{Row 1} \\ \text{Row 2}\end{array} & \begin{bmatrix} a & b \\ c & d \end{bmatrix}\end{array}$$

- Matrix Dimensions: (# of Rows) x (# of Columns)
- Matrix Addition and Subtraction

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \pm \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a \pm e & b \pm f \\ c \pm g & d \pm h \end{bmatrix}$$

Linear Algebra Review: Matrices

- Matrices Multiplication

$$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \times \begin{bmatrix} x & z & k \\ y & g & l \end{bmatrix} = \begin{bmatrix} ax+by & az+bg & ak+bl \\ cx+dy & cz+dg & ck+dl \\ ex+fy & ez+fg & ek+fl \end{bmatrix}$$

(3 x 2) x (2 x 3) = (3x3)

- Multiplying an (n x m) matrix by (m x k) matrix results in (n x k) matrix
- Matrix Transpose

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \qquad \mathbf{M}^T = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}$$

(3 x 2) (2 x 3)

Linear Algebra Review: Matrices

- Inverse of matrix $\mathbf{A}$ denoted by $\mathbf{A}^{-1}$

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad \text{where} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Identity matrix if $\mathbf{A}$ is (3 x 3)

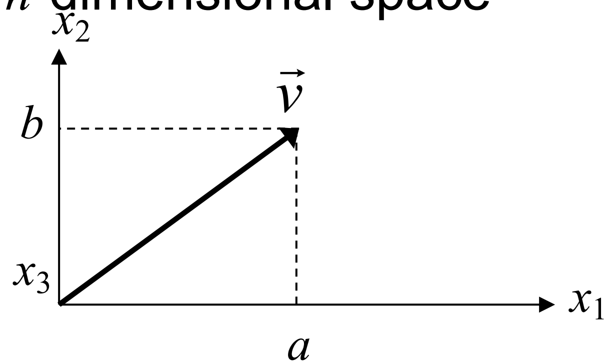
- For a (2 x 2) matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \det(\mathbf{A}) = ad - bc \qquad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

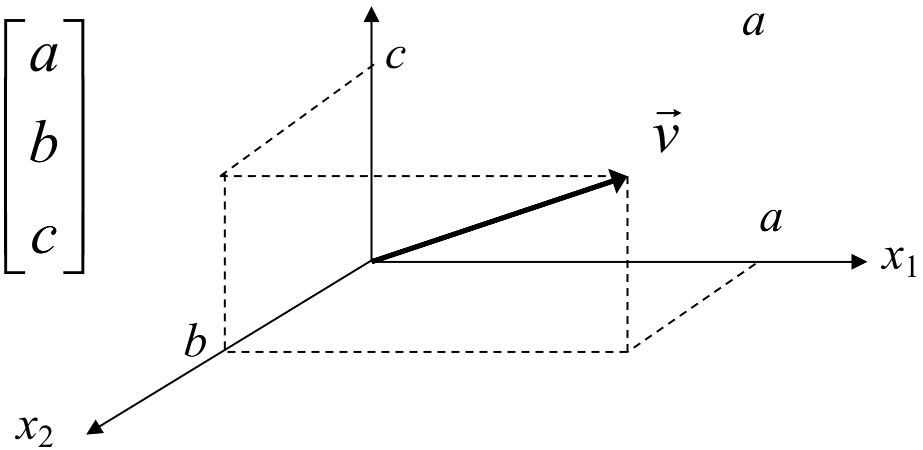
Linear Algebra Review: Vectors

- Vector = n x 1 matrix
- Represents a straight line in n-dimensional space

$$\vec{v} = \begin{bmatrix} a \\ b \end{bmatrix}$$



$$\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

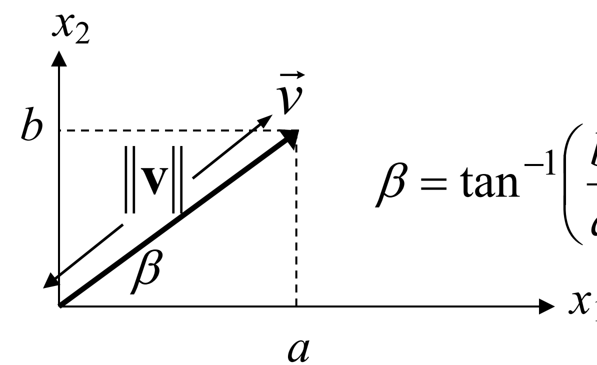


- A vector is denoted as $\vec{v}$ or $\mathbf{v}$

Linear Algebra Review: Vectors

- Vector Magnitude (Norm): Gives the length of a vector

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \|\mathbf{v}\| = \sqrt{a^2 + b^2}$$



$$\beta = \tan^{-1}\left(\frac{b}{a}\right)$$

- Unit Vector: Vector with norm = 1
- For an n-dimensional vector

$$\mathbf{v} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} a_i^2}$$

- The norm of the vector squared is equivalent to

$$\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1^2 + a_2^2 + \ldots + a_n^2 = \sum_{i=1}^{n} a_i^2$$

(1 x n) (n x 1) (1 x 1)

Linear Algebra Review: Vectors

- Vectors Dot Product

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} c \\ d \end{bmatrix}$$

$$\mathbf{v} \cdot \mathbf{u} = \mathbf{v}^T \mathbf{u} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = ac + bd$$

(Always (1 x 1))

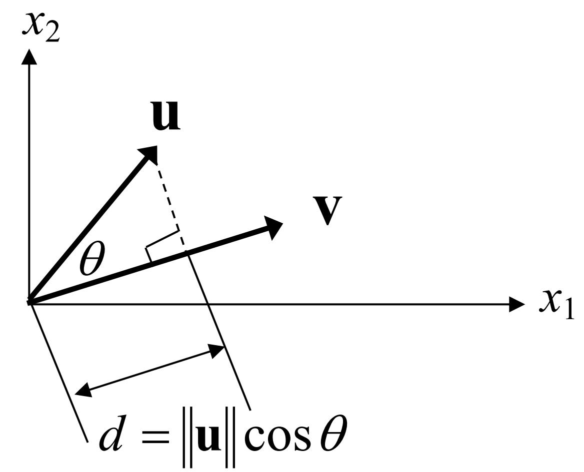
- For an n-dimensional vector

$$\mathbf{v} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \mathbf{u} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \mathbf{v} \cdot \mathbf{u} = \mathbf{v}^T \mathbf{u} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i$$

Linear Algebra Review: Vectors

- Dot product can be expressed as

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\|.\|\mathbf{u}\| \cos\theta$$



- $\|\mathbf{u}\| \cos\theta$ is the projection of the vector $\mathbf{u}$ on the vector $\mathbf{v}$
- If $\mathbf{v}$ is a unit vector, then the dot product is equivalent to the projection of the vector $\mathbf{u}$ on the vector $\mathbf{v}$
- If both vectors are unit vectors, the dot product will be maximum if both vectors are perfectly aligned
- If two vectors are orthogonal, the dot product will equal 0

Linear Algebra Review

- Matrix Calculus

$$\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}^T\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial(\mathbf{x}^T\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} \longrightarrow \frac{\partial \mathbf{x}^T\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

$$\frac{\partial(\mathbf{a}^T\mathbf{X}\mathbf{b})}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

Function Optimization

- To find the minimum (maximum) of a function, take the derivative and equate with zero

$$\min f(x) = f(x^*) \quad \text{where} \quad \frac{df(x)}{dx}\Big|_{x^*} = 0$$

- Example

$$f(x) = (x-2)^2$$

$$f'(x) = 2(x-2) = 2x - 4 = 0$$

$$x^* = 2$$

$$f(x^*) = 0$$