

# Preliminaries

Assignment Project Exam Help

<https://powcoder.com>

---

Add WeChat powcoder

# Who should take this class?

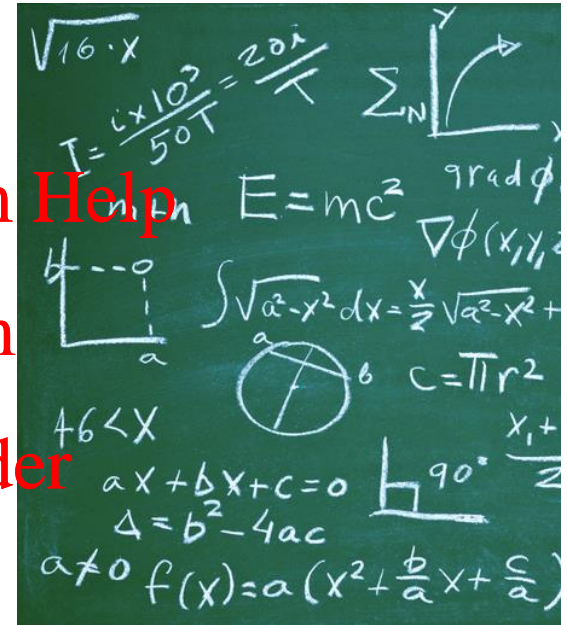
- This is a difficult, math- and programming-intensive class geared primarily towards graduate students

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- Historically, much fewer undergraduates manage an A than graduate students



# Course Prerequisites

- Linear algebra
- Multivariate Calculus, including partial derivatives
- Probability
- Comfort with programming in Python

Assignment Project Exam Help

<https://powcoder.com>

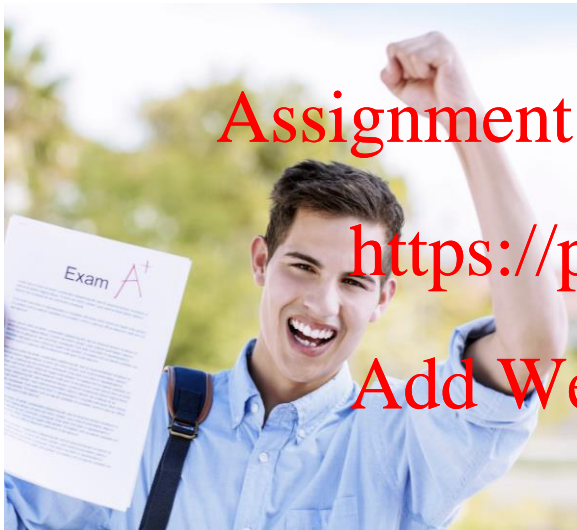
- [Fundamentals of Data Science \(CS 365\)](#) is a great prerequisite for this course
  - serves as a preparation including, but not limited, to the courses CS460, CS506, CS542 and CS565
- [Intro to Optimization \(CAS CS 507\)](#)
  - is not a formal prerequisite, but is highly recommended before taking this class

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Sufficient background

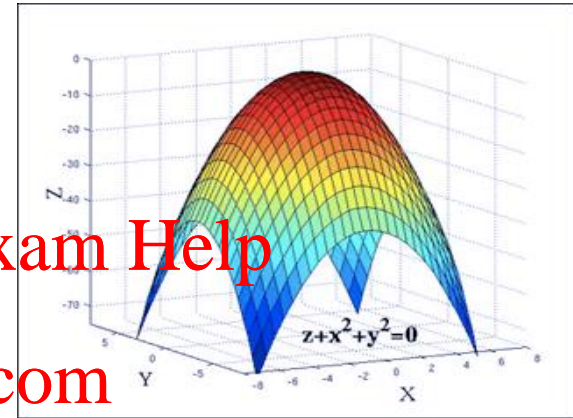


Insufficient background

# Course Prerequisites

- Multivariate Calculus

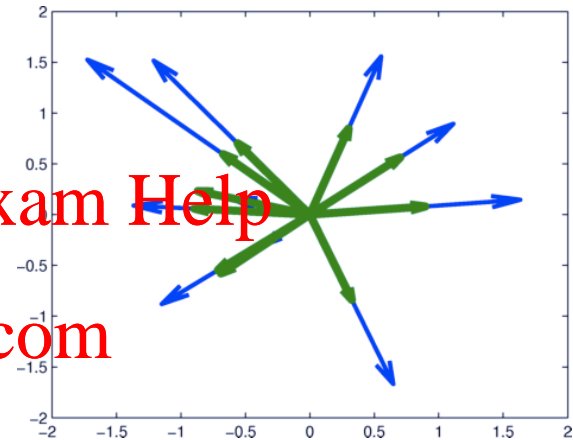
- Vectors; dot product
- Determinants; cross product
- Matrices; inverse matrices
- Square systems; equations of planes
- Parametric equations for lines and curves
- Max-min problems; least squares
- Second derivative test; boundaries and infinity
- Level curves; partial derivatives; tangent plane approximation
- Differentials; chain rule
- Gradient; directional derivative; tangent plane
- Lagrange multipliers
- Non-independent variables
- Double integrals
- Change of variables



- and other Calculus concepts such as convexity, etc.

# Course Prerequisites

- Linear algebra
  - Vectors and matrices
    - Basic Matrix Operations
    - Determinants, norms, trace
    - Special Matrices
  - Matrix inverse
  - Matrix rank
  - Eigenvalues and Eigenvectors
  - Matrix Calculus



Assignment Project Exam Help

<https://powcoder.com>

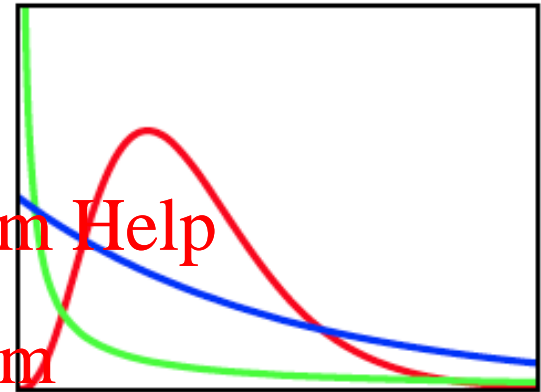
Add WeChat powcoder

# Course Prerequisites

- Probability

- Rules of probability, conditional probability and independence, Bayes rule

<https://powcoder.com>



- Random variables (expected value, variance, their properties); discrete and continuous variables, density functions, vector random variables, covariance, joint distributions
- Common distributions: Normal, Bernoulli, Binomial, Multinomial, Uniform, etc.

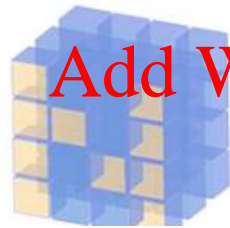
A review: <http://cs229.stanford.edu/section/cs229-prob.pdf>

# Course Prerequisites

Assignment Project Exam Help



<https://powcoder.com>



Add WeChat powcoder

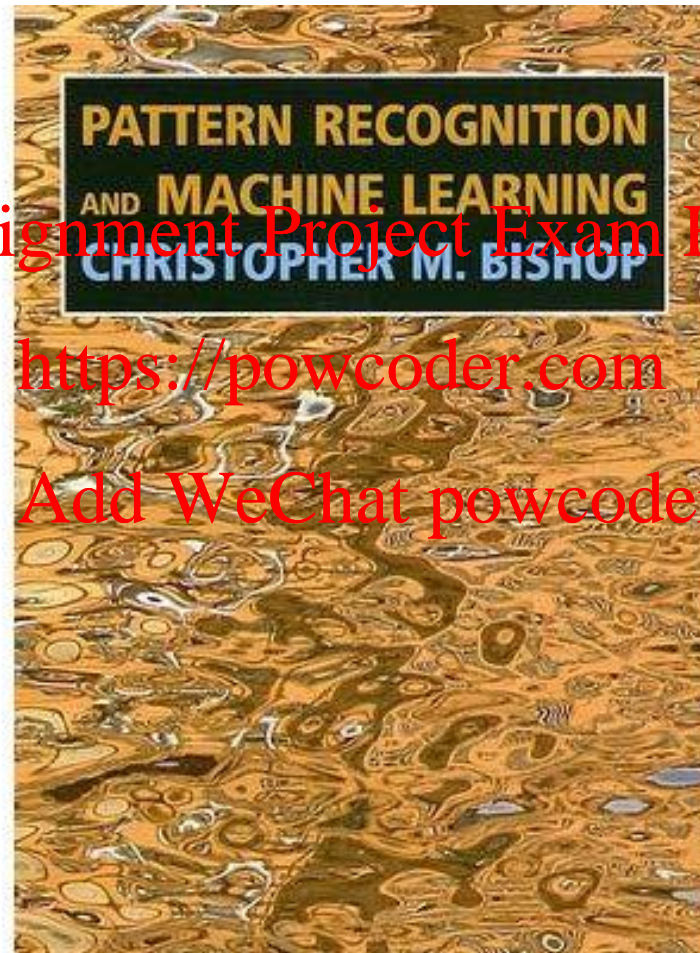
NumPy



*“..but I really want to take this course!”*

- If you lack any of these prerequisites, you SHOULD NOT take this class
- we cannot teach you the class material and also the prerequisite material  
*Assignment Project Exam Help*  
*<https://powcoder.com>*
- we are not miracle workers!  
*Add WeChat powcoder*
- instead, please consider these alternative courses:
  - *EC 414 Introduction to Machine Learning*
  - *CS 506 Computational Tools for Data*
  - *CS 504 Data Mechanics*

# Read the book



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors
- Matrix Calculus

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors
- Matrix Calculus

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Vector

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

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

Assignment Project Exam Help  
<https://powcoder.com>  
Add WeChat powcoder

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

$$\mathbf{v}^T = [v_1 \quad v_2 \quad \dots \quad v_n]$$

$T$  denotes the transpose operation

# Vector

- We'll default to column vectors in this class

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

Assignment Project Exam Help  
<https://powcoder.com>  
Add WeChat powcoder

# Matrix

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

Assignment Project Exam Help

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

- If  $m = n$ , we say that  $\mathbf{A}$  is square.

# Basic Matrix Operations

- What you should know:
  - Addition
  - Scaling
  - Dot product
  - Multiplication
  - Transpose
  - Inverse / pseudoinverse
  - Determinant / trace

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Vectors

- **Norm**

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}.$$

- More formally, a norm is any function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  that satisfies 4 properties.

Assignment Project Exam Help

<https://powcoder.com>

- **Non-negativity:** For all  $x \in \mathbb{R}^n$ ,  $f(x) \geq 0$
- **Definiteness:**  $f(x) = 0$  if and only if  $x = 0$ .
- **Homogeneity:** For all  $x \in \mathbb{R}^n$ ,  $t \in \mathbb{R}$ ,  $f(tx) = |t|f(x)$
- **Triangle inequality:** For all

$$x, y \in \mathbb{R}^n, f(x + y) \leq f(x) + f(y)$$

# Matrix Operations

- **Example Norms**

$$\|x\|_1 = \sum_{i=1}^n |x_i| \qquad \|x\|_\infty = \max_i |x_i|$$

Assignment Project Exam Help

- General  $\ell_p$  norms:

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

<https://powcoder.com>  
Add WeChat powcoder

# Matrix Operations

- Inner product (dot product) of vectors
  - Multiply corresponding entries of two vectors and add up the result
  - $\mathbf{x} \cdot \mathbf{y}$  is also  $|\mathbf{x}| |\mathbf{y}| \cos(\theta)$  (the angle between  $\mathbf{x}$  and  $\mathbf{y}$ )

Assignment Project Exam Help

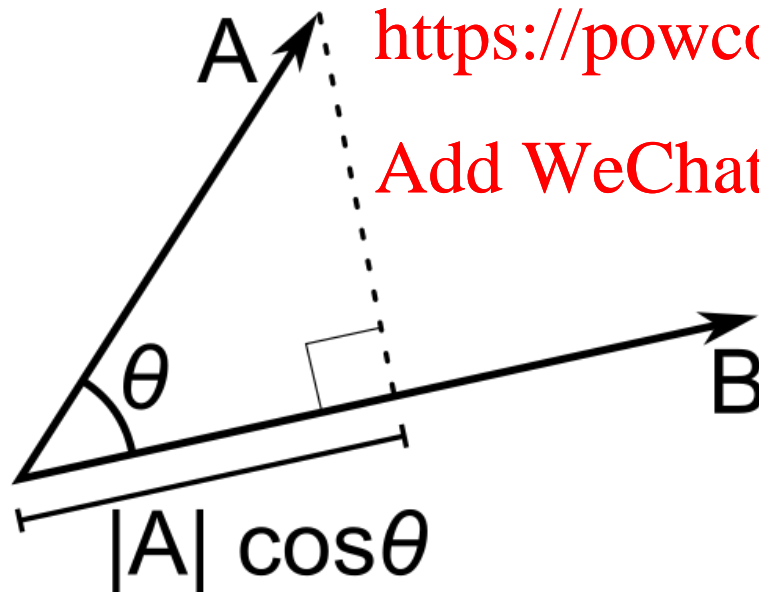
<https://powcoder.com>

Add WeChat powcoder

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

# Matrix Operations

- Inner product (dot product) of vectors
  - If B is a unit vector, then  $A \cdot B$  gives the length of A which lies in the direction of B



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Operations

- The product of two matrices

Matrix multiplication is associative:  $(AB)C = A(BC)$ .

Matrix multiplication is distributive:  $A(B + C) = AB + AC$ .

Matrix multiplication is, in general, *not* commutative; that is, it can be the case that  $AB \neq BA$ . (For example, if  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{n \times q}$ , the matrix product  $BA$  does not even exist if  $m$  and  $q$  are not equal.)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Operations

- Powers

- By convention, we can refer to the matrix product  $AA$  as  $A^2$ , and  $AAA$  as  $A^3$ , etc.
- Obviously only square matrices can be multiplied that way

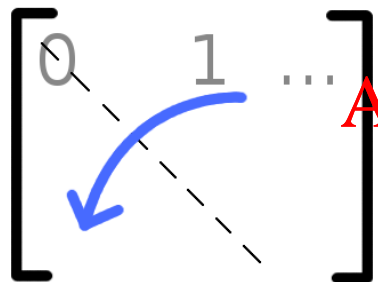
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Operations

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



Assignment Project Exam Help  
<https://powcoder.com>

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

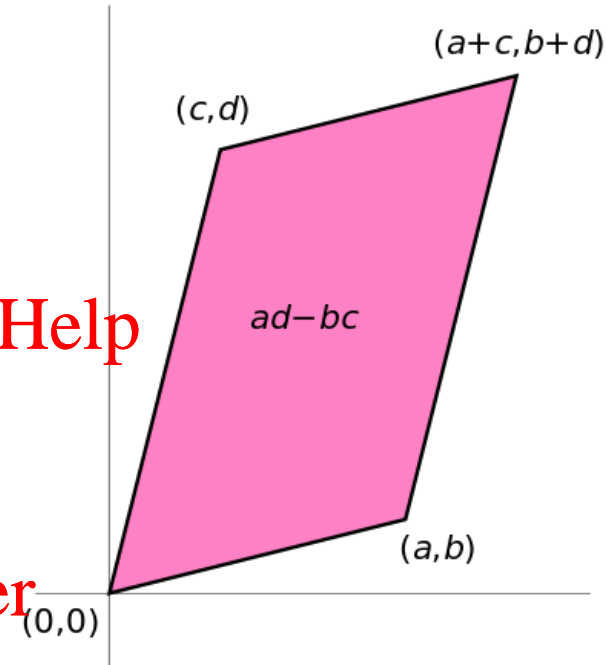
- A useful identity: Add WeChat powcoder

$$(ABC)^T = C^T B^T A^T$$

# Matrix Operations

- Determinant

- $\det(\mathbf{A})$  returns a scalar
- Represents area (or volume) of the parallelogram described by the vectors in the rows of the matrix



- For  $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ,  $\det(\mathbf{A}) = ad - bc$

- Properties:  $\det(\mathbf{AB}) = \det(\mathbf{BA})$

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A})$$

$$\det(\mathbf{A}) = 0 \Leftrightarrow \mathbf{A} \text{ is singular}$$



# Matrix Operations

- Trace

$\text{tr}(\mathbf{A})$  = sum of diagonal elements

$$\text{tr}\left(\begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix}\right) = 1 + 7 = 8$$

Assignment Project Exam Help

– Invariant to a lot of transformations, so it's used sometimes in proofs. (Rarely in this class though.)

Add WeChat powcoder

– Properties:

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

# Matrix Operations

- **Vector Norms**

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad \|x\|_\infty = \max_i |x_i|.$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad \|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

- Matrix norms: Norms can also be defined for matrices, such as

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\text{tr}(A^T A)}.$$

# Special Matrices

- Symmetric matrix

$$\mathbf{A}^T = \mathbf{A}$$

- Skew-symmetric matrix

<https://powcoder.com>

$$\mathbf{A}^T = -\mathbf{A}$$

Add WeChat powcoder

- Identity matrix  $\mathbf{I}$

- Diagonal matrix

$$\begin{bmatrix} 1 & 2 & 5 \\ 2 & 1 & 7 \\ 5 & 7 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -2 & -5 \\ 2 & 0 & -7 \\ 5 & 7 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}$$

# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors
- Matrix Calculate

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Inverse

- Given a matrix  $\mathbf{A}$ , its inverse  $\mathbf{A}^{-1}$  is a matrix such that  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
- E.g.  $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$   
<https://powcoder.com>
- Inverse does not always exist. If  $\mathbf{A}^{-1}$  exists,  $\mathbf{A}$  is *invertible* or *non-singular*. Otherwise, it's *singular*.
- Useful identities, for matrices that are invertible:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$\mathbf{A}^{-T} \triangleq (\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$

# Matrix Operations

- Pseudoinverse

- Say you have the matrix equation  $AX=B$ , where A and B are known, and you want to solve for X
- You could calculate the inverse and pre-multiply by it:  
 $A^{-1}AX=A^{-1}B \rightarrow X=A^{-1}B$
- Python command would be `np.linalg.inv(A)*B`
- But calculating the inverse for large matrices often brings problems with computer floating-point resolution (because it involves working with very small and very large numbers together).
- Or, your matrix might not even have an inverse.

# Matrix Operations

- Pseudoinverse

- Fortunately, there are workarounds to solve  $AX=B$  in these situations. And python can do them!
- Instead of taking an inverse, directly ask python to solve for  $X$  in  $AX=B$ , by typing `np.linalg.solve(A, B)`
- Python will try several appropriate numerical methods (including the pseudoinverse if the inverse doesn't exist)
- Python will return the value of  $X$  which solves the equation
  - If there is no exact solution, it will return the closest one
  - If there are many solutions, it will return the smallest one

# Matrix Operations

- Python example:

$$AX = B$$

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

```
>> import numpy as np
>> x = np.linalg.solve(A,B)
x =
    1.0000
   -0.5000
```



# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors
- Matrix Calculate

Assignment Project Exam Help

<https://powcoder.com>

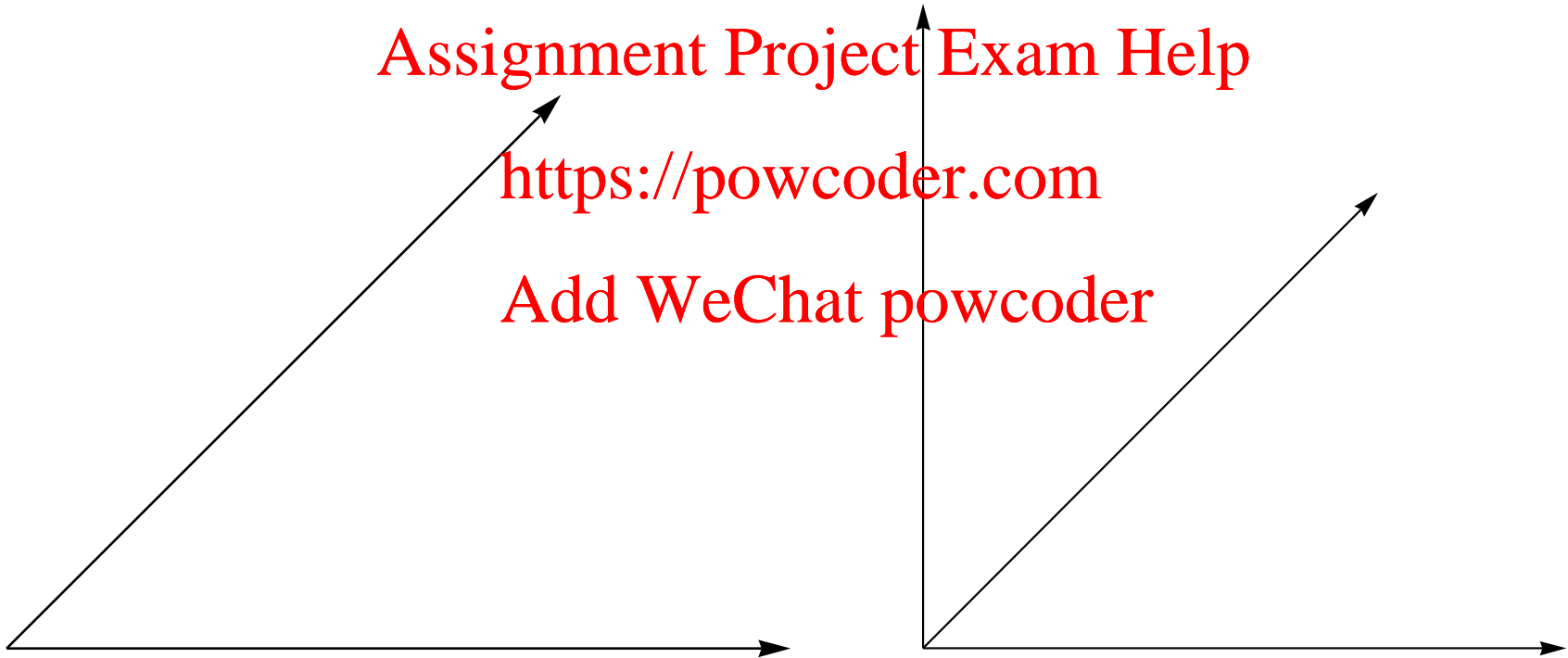
Add WeChat powcoder

# Linear independence

- Suppose we have a set of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$
- If we can express  $\mathbf{v}_1$  as a linear combination of the other vectors  $\mathbf{v}_2, \dots, \mathbf{v}_n$ , 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  $\mathbf{v}_2, \dots, \mathbf{v}_n$ . (E.g.  $\mathbf{v}_1 = .7 \mathbf{v}_2 - .7 \mathbf{v}_4$ )
- If no vector is linearly dependent on the rest of the set, the set is linearly *independent*.
  - Common case: a set of vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  is always linearly independent if each vector is perpendicular to every other vector (and non-zero)

# Linear independence

Linearly independent set      Not linearly independent



# Matrix rank

- Column/row rank

$\text{col-rank}(\mathbf{A}) =$  the maximum number of linearly independent column vectors of  $\mathbf{A}$

$\text{row-rank}(\mathbf{A}) =$  the maximum number of linearly independent row vectors of  $\mathbf{A}$

– Column rank always equals row rank

<https://powcoder.com>  
Add WeChat powcoder

- Matrix rank

$$\text{rank}(\mathbf{A}) \triangleq \text{col-rank}(\mathbf{A}) = \text{row-rank}(\mathbf{A})$$

# Matrix rank

- For transformation matrices, the rank tells you the dimensions of the output
- E.g. if rank of **A** is 1, then the transformation

Assignment Project Exam Help

$$\mathbf{p}' = \mathbf{A}\mathbf{p}$$

<https://powcoder.com>

maps points onto a line.

Add WeChat powcoder

- Here's a matrix with rank 1:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + y \\ 2x + 2y \end{bmatrix}$$

← All points get mapped to the line  $y=2x$

# Matrix rank

- If an  $m \times m$  matrix is rank  $m$ , we say it's "full rank"
  - Maps an  $m \times 1$  vector uniquely to another  $m \times 1$  vector
  - An inverse matrix can be found
- If rank  $< m$ , we say it's "singular"
  - At least one dimension is getting collapsed. No way to look at the result and tell what the input was
  - Inverse does not exist
- Inverse also doesn't exist for non-square matrices

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors(SVD)
- Matrix Calculus

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Eigenvector and Eigenvalue

- An eigenvector  $\mathbf{x}$  of a linear transformation  $A$  is a non-zero vector that, when  $A$  is applied to it, does not change direction. <https://powcoder.com>

Add WeChat powcoder

$$Ax = \lambda x, \quad x \neq 0.$$



# Eigenvector and Eigenvalue

- An eigenvector  $\mathbf{x}$  of a linear transformation  $A$  is a non-zero vector that, when  $A$  is applied to it, does not change direction. <https://powcoder.com>
- Applying  $A$  to the eigenvector only scales the eigenvector by the scalar value  $\lambda$ , called an eigenvalue.

$$Ax = \lambda x, \quad x \neq 0.$$

# Properties of eigenvalues

- The trace of a  $A$  is equal to the sum of its eigenvalues:

$$\text{tr}A = \sum_{i=1}^n \lambda_i.$$

Assignment Project Exam Help

- The determinant of  $A$  is equal to the product of its eigenvalues

<https://powcoder.com>

$$|A| = \prod_{i=1}^n \lambda_i.$$

Add WeChat powcoder

- The rank of  $A$  is equal to the number of non-zero eigenvalues of  $A$ .
- The eigenvalues of a diagonal matrix  $D = \text{diag}(d_1, \dots, d_n)$  are just the diagonal entries  $d_1, \dots, d_n$

# Diagonalization

- Eigenvalue equation:

$$AV = VD$$

Assignment Project Exam Help

$$A = VDV^{-1}$$

<https://powcoder.com>

- Where D is a diagonal matrix of the eigenvalues

Add WeChat powcoder

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

# Diagonalization

- Eigenvalue equation:

$$AV = VD$$

Assignment Project Exam Help

$$A = VDV^{-1}$$

<https://powcoder.com>

- Assuming all  $\lambda_i$ 's are unique:

Add WeChat powcoder

$$A = VDV^T$$

- Remember that the inverse of an orthogonal matrix is just its transpose and the eigenvectors are orthogonal

# Symmetric matrices

- Properties:
  - For a symmetric matrix  $A$ , all the eigenvalues are real.
  - The eigenvectors of  $A$  are orthonormal.

Add WeChat powcoder

$$A = V D V^T$$

# Symmetric matrices

- Therefore:

$$x^T A x = x^T V D V^T x = y^T D y = \sum_{i=1}^n \lambda_i y_i^2$$

– where  $y = V^T x$

- So, if we wanted to find the vector  $x$  that:

$$\max_{x \in \mathbb{R}^n} x^T A x \quad \text{subject to } \|x\|_2^2 = 1$$

# Symmetric matrices

- Therefore:

$$x^T A x = \sum_{i=1}^n \lambda_i y_i^2$$

– where  $y = V^T x$

- So, if we wanted to find the vector  $x$  that:

$$\max_{x \in \mathbb{R}^n} x^T A x \quad \text{subject to } \|x\|_2^2 = 1$$

– Is the same as finding the eigenvector that corresponds to the largest eigenvalue.

# Matrix Algebra Review

- Vectors and matrices
  - Basic Matrix Operations
  - Determinants, norms, trace
  - Special Matrices
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors(SVD)
- Matrix Calculus

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Matrix Calculus – The Gradient

- Let a function  $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$  take as input a matrix  $A$  of size  $m \times n$  and returns a real value.

Assignment Project Exam Help

- Then the **gradient** of **f**:

<https://powcoder.com>

Add WeChat powcoder

$$\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \dots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \dots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \dots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

# Matrix Calculus – The Gradient

- Every entry in the matrix is:  $(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$ .
- the size of  $\nabla_A f(A)$  is always the same as the size of A. So if A is just a vector x:

<https://powcoder.com>

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

# Exercise

- Example:

For  $x \in \mathbb{R}^n$ , let  $f(x) = b^T x$  for some known vector  $b \in \mathbb{R}^n$

Assignment Project Exam Help

<https://powcoder.com>

$$f(x) = [b_1 \quad b_2 \quad \dots \quad b_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Add WeChat powcoder

- Find:  $\frac{\partial f(x)}{\partial x_k} = ?$

$$\nabla_x f(x) = ?$$

# Exercise

- Example:

For  $x \in \mathbb{R}^n$ , let  $f(x) = b^T x$  for some known vector  $b \in \mathbb{R}^n$

Assignment Project Exam Help

$$f(x) = \sum_{i=1}^n b_i x_i$$

<https://powcoder.com>

Add WeChat powcoder

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n b_i x_i = b_k.$$

- From this we can conclude that:  $\nabla_x b^T x = b$ .

# Matrix Calculus – The Gradient

- Properties

- $\nabla_x (f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x).$

- For  $t \in \mathbb{R}$ ,  $\nabla_x (t f(x)) = t \nabla_x f(x).$

Add WeChat powcoder

# Matrix Calculus – The Jacobian

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

# Matrix Calculus – The Hessian

- The Hessian matrix with respect to  $x$ , written  $\nabla_x^2 f(x)$  or simply as  $H$  is the  $n \times n$  matrix of partial derivatives

<https://powcoder.com>  
Add WeChat powcoder

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

# Matrix Calculus – The Hessian

- Each entry can be written as:  $\nabla_x^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ .

Assignment Project Exam Help

- Exercise: Why is the Hessian always symmetric?

<https://powcoder.com>

Add WeChat powcoder



# Matrix Calculus – The Hessian

- Each entry can be written as:

$$\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}.$$

- The Hessian is always symmetric, because

<https://powcoder.com>

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

- This is known as Schwarz's theorem: The order of partial derivatives don't matter as long as the second derivative exists and is continuous.

# Matrix Calculus – The Hessian

- Note that the hessian is not the gradient of whole gradient of a vector (this is not defined). It is actually the gradient of **every entry** of the gradient of the vector.

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

# Matrix Calculus – The Hessian

- Eg, the first column is the gradient of  $\frac{\partial f(x)}{\partial x_1}$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder.

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

# Common vector derivatives

Scalar derivative		Vector derivative	
$f(x)$	$\rightarrow \frac{df}{dx}$	$f(\mathbf{x})$	$\rightarrow \frac{df}{d\mathbf{x}}$
$bx$	$\rightarrow b$	$\mathbf{x}^T \mathbf{B}$	$\rightarrow \mathbf{B}$
$bx$	$\rightarrow b$	$\mathbf{x}^T \mathbf{b}$	$\rightarrow \mathbf{b}$
$x^2$	$\rightarrow 2x$	$\mathbf{x}^T \mathbf{x}$	$\rightarrow 2\mathbf{x}$
$bx^2$	$\rightarrow 2bx$	$\mathbf{x}^T \mathbf{B} \mathbf{x}$	$\rightarrow 2\mathbf{B} \mathbf{x}$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# PSet 1 Out Today

- Due in 1 week: 9/15 11:59pm GMT -5 (Boston Time)

Assignment Project Exam Help

- Diagnostic homework covering topics covered in prereqs <https://powcoder.com>  
Add WeChat powcoder

- Additional examples in lab this week (Group A for in-person lab rotations)

# Next Class

## Supervised Learning I: Regression:

regression, linear hypothesis, SSD cost; gradient descent; normal equations; maximum likelihood;

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

**Reading:** Bishop 1.2-1.2.4, 3.1-3.1.1