
Linear Data Analysis

Offered as: CISC 271, Winter 2024

RANDY E ELLIS, JANUARY 2024

Contents

Contents	i
1 Introduction, With a Review of Eigenvalues and Eigenvectors	1
1.1 Eigenvalues and Eigenvectors	4
1.2 Characteristic Polynomials and Characteristic Equations	5
1.3 Eigenfacts: Matrix Properties and Eigenvalues	5
1.4 Eigenvectors	7
1.5 Extra Notes: Extensions to Linear Algebra	7
1.6 Extra Notes: Characterization of a Real Symmetric Matrix	9
2 Graphs: The Adjacency Matrix	10
2.1 Relevant Definitions	12
2.2 Adjacency Matrix of a Graph	13
2.3 Extra Notes: Incidence Matrix of a Graph	17
3 Graphs: The Laplacian Matrix	19
3.1 Eigenvalues of a Laplacian Matrix	21
3.2 The Fiedler Vector of a Graph	24
3.3 Extra Notes: Properties of a Laplacian Matrix	25
4 Vector Spaces	26
4.1 Vector Space: Properties	26
4.1.1 Vector Space: Interpretation	27
4.1.2 Vector Space: Examples	28
4.2 Matrices and Vector Spaces	30
4.3 Echelon Forms of a Matrix $A \in \mathbb{R}^{m \times n}$	31
4.4 Null Space and the RREF	33
5 Spanning Sets and Basis Vectors	38
5.1 Orthogonal Basis and Orthonormal Basis	40
5.2 Rank-Nullity Theorem	41
5.3 Orthogonal Subspaces	43

6	Diagonalizable Matrices	44
6.1	Similar Matrices	44
6.2	Eigenvectors as a Basis	45
6.3	Eigenvector Basis	46
6.4	Nondiagonalizable Matrices	47
6.5	Matrix Powers	48
6.6	Extra Notes: Small Perturbations	49
7	Normal Matrices and Spectral Decomposition	52
7.1	Orthogonal Matrices	52
7.2	Symmetric Matrices	53
7.3	Normal Matrices	55
7.4	Skew-Symmetric Matrices	56
8	Positive [Semi-]Definite Matrices	58
8.1	Quadratic Form of a Symmetric Matrix	59
8.2	Example: Product of a Full-Rank Matrix and its Transpose	59
8.3	Statistics of Vectors: Means and Variance	60
8.4	Example: Covariance Matrix in Statistics	62
8.5	Example: Linear Elastic Structures	63
9	Design Matrix and Standardized Data	65
9.1	Data Matrix and Design Matrix	67
9.2	Standardized Data or Z Score	68
10	Review #1: Graphs and Diagonalizable Matrices	71
11	Orthogonal Projection	72
11.1	Projecting to a 1D Subspace: Vector to Vector	72
11.2	Projecting to a 2D Subspace: Vector to Basis	74

12	Patterns – Linear Regression	77
12.1	Residual Error	78
12.2	Linear Regression	79
12.3	Example: Hooke’s Law	79
12.4	Linear Regression – One Dependent Variable, Plus Intercept	80
12.5	Data Matrix For Linear Regression	81
12.6	Assessment Of Linear Regression	83
12.7	Extra Notes For Derivations of Linear Regression	84
13	Cross-Validating Linear Regression	87
13.1	Training and Testing	87
13.2	Linear Regression and the Design Matrix X	88
13.3	Measuring Error in Linear Regression	88
13.4	Cross-Validation of Linear Regression: Leaving Data Out	89
13.5	Cross-Validation of Linear Regression: K-Fold Analysis	90
13.6	Cross-Validation of Linear Regression: Monte Carlo Methods	91
13.7	Example: 13 Data With 2 Outliers	91
14	SVD – Singular Value Decomposition	94
14.1	Eigenvectors of the matrix $[A^T A]$	94
14.2	Eigenvectors of the matrix $[A A^T]$	95
14.3	The Singular Value Decomposition, or SVD	96
14.4	Using the SVD	97
15	Orthonormal Basis Vectors and the SVD	98
15.1	SVD of a Square Matrix	98
15.2	SVD of a Non-Square Matrix	100
15.3	The SVD as an Approximate Basis for a Vector Space	101
15.4	Some SVD Properties	102
16	Review #2: Matrices and Linear Regression	104

17 Matrix Approximation	105
17.1 Matrix Norms	105
17.1.1 Eigenvalues and Singular Values	106
17.2 A Matrix As A Series	107
17.3 Matrix Approximation: Eckart-Young Theorem	108
17.4 Approximations and The Scree Plot	109
18 Principal Components Analysis – PCA	111
18.1 Motivation, by Example	111
18.2 Zero-Mean Data Matrix	112
18.3 Principal Components Analysis as an SVD	113
18.4 Using the SVD to Compute PCA Scores	114
19 Unsupervised Learning – K-Means Clustering	117
19.1 Supervised and Unsupervised Classification	117
19.2 Unsupervised Classification: Data Clustering	117
19.3 Clustering – Iris Data	118
19.4 K-Means Clustering	119
19.4.1 A K-Means Algorithm	120
19.4.2 Example: Fisher’s Iris Data	121
20 Classification – Linear Separability	124
20.1 Separation of Two Clusters	124
20.2 Separation of Three or More Clusters	127
21 PCA – Matrix Algebra and Dimensionality Reduction	130
21.1 Scatter Matrix, SVD, and PCA	131
21.2 PCA and Low-Rank Approximation	132
21.3 PCA for Dimensionality Reduction	133
22 Review #3: SVD, PCA, And Dimensionality Reduction	134

23	PCA and the Rayleigh Quotient	135
23.1	Zero-Mean Data and Maximization	135
23.2	Rayleigh Quotient of a Symmetric Matrix	136
24	Patterns – Linear Discriminant Analysis, or LDA	138
24.1	Scatter Matrices For Labelled Data	139
24.2	Fisher’s Linear Discriminant	141
24.3	LDA Using Test Data	142
24.4	Extra Notes – Maximum of Rayleigh Quotient	143
24.5	Extra Notes – Fisher’s Linear Discriminant	144
25	Classification – Assessment With Confusion Matrix	146
25.1	Relative Confusion Matrix	148
26	Classification – Assessment With ROC Curve	150
26.1	Receiver Operator Characteristic – ROC	150
26.2	Example: Two Variants of a Virus	152
27	Classification – Single Artificial Neuron	156
27.1	Hyperplane For An Artificial Neuron	158
28	Review #4: LDA, Assessment, and Odds of Occurrence	160
29	Odds of Occurrence and Probability	161
29.1	Logistic Function – Some Properties	165
30	Supervised Learning – Perceptron Rule	167
30.1	The Perceptron Rule	167
30.2	Perceptron Rule – Batch Learning	170
30.3	Example: Performance on the Iris Data Set	170

31	Classification – Logistic Regression	172
31.1	Semilinear Activation – Logistic Function As A Sigmoid Curve	173
31.2	Models Of Residual Error	174
31.3	Implementations Of Logistic Activation	176
31.3.1	Example – Fisher’s Iris Data	176
32	Nonlinear Separation – Embeddings and Gram Matrix	179
32.1	Linear Separation Using an Embedding	180
32.2	Kernel Functions and the Gram Matrix	182
32.3	Some Kernel Functions for Row Spaces	183
32.4	Extra Notes for The Gram Matrix and Kernel Functions	185
33	Nonlinear Separation – Kernel PCA	186
33.1	Principal Components Analysis And Scatter Of Observations	186
33.2	Kernel Functions and the Gram Matrix	189
33.3	Example – Kernel PCA For Fisher’s Iris Data	190
34	Review #5: Machine Learning	192
35	Spectral Clustering Of Data	193
35.1	Multiple Clusters In Graphs	193
35.1.1	Example: 30 Vertices In 3 Clusters	194
35.2	Graph Representation Of Simple Data	195
35.3	Weighted Graphs	199
35.4	Extra Notes on The Laplacian Matrix	199
36	The Curse of Dimensionality	202
36.1	Exploring Dimensionality – Hypercubes	202
36.2	Exploring Dimensionality – Hypercube Labels	203
36.3	Exploring Dimensionality – Uniformly Distributed Data	205
36.4	Exploring Dimensionality – Normally Distributed Data	206
36.5	Exploring Dimensionality – Summary	211
37	Course Summary – Linear Data Analysis	212