

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: Adjacency Matrix and Laplacian Matrix	10
2.1 Relevant Definitions	12
2.2 Adjacency Matrix of a Graph	13
2.3 Weighted Adjacency Matrix of a Graph	17
2.4 Laplacian Matrix of a Graph	17
2.5 Eigenvalues of a Laplacian Matrix	20
2.6 The Fiedler Vector of a Graph	22
2.7 Extra Notes: Incidence Matrix of a Graph	23
2.8 Extra Notes: Properties of a Laplacian Matrix	25
3 Vector Spaces	26
3.1 Vector Space: Properties	26
3.1.1 Vector Space: Interpretation	27
3.1.2 Vector Space: Examples	28
3.2 Matrices and Vector Spaces	30
3.3 Echelon Forms of a Matrix $A \in \mathbb{R}^{m \times n}$	31
3.4 Null Space and the RREF	33

4	Spanning Sets and Basis Vectors	38
4.1	Orthogonal Basis and Orthonormal Basis	40
4.2	Rank-Nullity Theorem	41
4.3	Orthogonal Subspaces	43
5	Diagonalizable Matrices	44
5.1	Similar Matrices	44
5.2	Eigenvectors as a Basis	45
5.3	Eigenvector Basis	46
5.4	Orthogonal Matrices	46
5.5	Symmetric Matrices	48
5.6	Skew-Symmetric Matrices	49
5.7	Nondiagonalizable Matrices	50
5.8	Extra Notes: Matrix Powers	52
5.9	Extra Notes: Small Perturbations	53
6	Spectral Decomposition and Positive [Semi–]Definite Matrices	55
6.1	The Spectral Decomposition	55
6.2	Positive-Definite and Positive-Semidefinite Matrices	57
6.3	Quadratic Form of a Symmetric Matrix	57
6.4	Example: Product of a Full-Rank Matrix and its Transpose	58
6.5	Statistics of Vectors: Means and Variance	59
6.6	Example: Covariance Matrix in Statistics	60
6.7	Example: Linear Elastic Structures	62
7	Design Matrix and Standardized Data	64
7.1	Data Matrix and Design Matrix	66
7.2	Standardized Data or Z Score	67
8	Orthogonal Projection	70
8.1	Projecting to a 1D Subspace: Vector to Vector	70
8.2	Projecting to a 2D Subspace: Vector to Basis	72

9	Review For Quiz #1: Graphs and Spectral Decomposition	75
10	Patterns – Linear Regression	76
10.1	Residual Error	77
10.2	Linear Regression	78
10.3	Example: Hooke’s Law	78
10.4	Linear Regression – One Dependent Variable, Plus Intercept	79
10.5	Data Matrix For Linear Regression	80
10.6	Assessment Of Linear Regression	82
10.7	Extra Notes For Derivations of Linear Regression	83
11	Cross-Validating Linear Regression	86
11.1	Training and Testing	86
11.2	Linear Regression and the Design Matrix \mathbf{X}	87
11.3	Measuring Error in Linear Regression	87
11.4	Cross-Validation of Linear Regression: Leaving Data Out	88
11.5	Cross-Validation of Linear Regression: K-Fold Analysis	89
11.6	Cross-Validation of Linear Regression: Monte Carlo Methods	90
11.7	Example: 13 Data With 2 Outliers	90
12	Review For Test #1: Basic Linear Analysis	93
13	SVD – Singular Value Decomposition	95
13.1	Eigenvectors of the matrix $[A^T A]$	95
13.2	Eigenvectors of the matrix $[AA^T]$	96
13.3	The Singular Value Decomposition, or SVD	97
13.4	Using the SVD	98
14	Orthonormal Basis Vectors and the SVD	99
14.1	SVD of a Square Matrix	99
14.2	SVD of a Non-Square Matrix	101
14.3	The SVD as an Approximate Basis for a Vector Space	102
14.4	Some SVD Properties	103

15 Review for Quiz #2: Matrices and Linear Regression	105
16 Principal Components Analysis – PCA	106
16.1 Motivation, by Example	106
16.2 Zero-Mean Data Matrix	107
16.3 Principal Components Analysis as an SVD	108
16.4 Using the SVD to Compute PCA Scores	109
16.5 Matrix Norms	111
16.6 Eigenvalues and Singular Values	112
16.7 Extra Notes: A Matrix As A Series	113
16.8 Extra Notes: Eckart-Young Theorem	114
17 PCA – Matrix Algebra and Dimensionality Reduction	116
17.1 Scatter Matrix, SVD, and PCA	117
17.2 PCA and Low-Rank Approximation	118
17.3 PCA for Dimensionality Reduction	119
17.4 Approximations and The Scree Plot	119
18 Review for Test #2: Matrices and Linear Regression	121
19 Unsupervised Learning – K-Means Clustering	123
19.1 Supervised and Unsupervised Classification	123
19.2 Unsupervised Classification: Data Clustering	123
19.3 Clustering – Iris Data	124
19.4 K-Means Clustering	125
19.4.1 A K-Means Algorithm	125
19.4.2 Example: Fisher’s Iris Data	126
20 Classification – Linear Separability	130
20.1 Separation of Two Clusters	130
20.2 Separation of Three or More Clusters	133
20.3 Hyperplane Specification With a Unit Normal Vector	136

21 Review for Quiz #3: SVD, PCA, And Dimensionality Reduction	137
22 Classification – Assessment With Confusion Matrix	138
22.1 Relative Confusion Matrix	140
23 Classification – Assessment With ROC Curve	142
23.1 Receiver Operator Characteristic – ROC	142
23.2 Example: Two Variants of a Virus	144
24 Review for Test #3: SVD, PCA, And Dimensionality Reduction	148
25 Odds of Occurrence and Probability	150
25.1 Logistic Function – Some Properties	154
25.2 Hyperplane Classification and Pseudo-Distance	155
26 Classification – Single Artificial Neuron	156
26.1 Data Matrix and Label Vector	157
26.1.1 Labels of Data for Artificial Neurons	157
26.2 Activation Function	158
26.3 Hyperplane For An Artificial Neuron	158
27 Review #4: LDA, Assessment, and Odds of Occurrence	160
28 Classification – Learning Weights and Logistic Regression	161
28.1 History – The Perceptron Rule	161
28.2 Logistic Activation of an Artificial Neuron	161
28.3 The Algorithm of Steepest Descent	163
28.4 Models Of Residual Error	166
28.5 Implementations Of Logistic Activation	168
28.5.1 Example – Fisher’s Iris Data	168
28.6 Extra Notes: Gradient of Squared-Error Objective	170

29 Nonlinear Separation – Embeddings and Gram Matrix	172
29.1 Linear Separation Using an Embedding	173
29.2 Kernel Functions and the Gram Matrix	175
29.3 Some Kernel Functions for Row Spaces	176
29.4 Extra Notes for The Gram Matrix and Kernel Functions	178
30 Review for Test #4: Separability, Assessment, and Odds of Occurrence	179
31 Nonlinear Separation – Kernel PCA	181
31.1 Principal Components Analysis And Scatter Of Observations	181
31.2 Kernel Functions and the Gram Matrix	184
31.3 Example – Kernel PCA For Fisher’s Iris Data	185
32 Spectral Clustering Of Data	187
32.1 Multiple Clusters In Graphs	187
32.1.1 Example: 30 Vertices In 3 Clusters	188
32.2 Graph Representation Of Simple Data	189
32.3 Weighted Graphs	193
32.4 Extra Notes on The Laplacian Matrix	193
33 Review for Quiz #5: Machine Learning	196
34 The Curse of Dimensionality	197
34.1 Exploring Dimensionality – Hypercubes	197
34.2 Exploring Dimensionality – Hypercube Labels	198
34.3 Exploring Dimensionality – Uniformly Distributed Data	200
34.4 Exploring Dimensionality – Normally Distributed Data	201
34.5 Exploring Dimensionality – Summary	206
35 Course Summary – Linear Data Analysis	207
36 Review for Test #5: Machine Learning	208
References	210