Contents

Contents i

1 Introduction 1

2 Eigenvalues and Eigenvectors 2

2.1 Eigenvalues and Eigenvectors 4

2.2 Characteristic Polynomials and Characteristic Equations 4

2.3 Eigenfacts: Matrix Properties and Eigenvalues 6

2.4 Eigenvectors 7

2.5 Extra Notes: Extensions to Linear Algebra 7

2.6 Extra Notes: Characterization of a Real Symmetric Matrix 9

3 Graphs: Adjacency Matrix and Laplacian Matrix 10

3.1 Relevant Definitions 12

3.2 Adjacency Matrix of a Graph 13

3.3 Weighted Adjacency Matrix of a Graph 17

3.4 Laplacian Matrix of a Graph 17

3.5 Eigenvalues of a Laplacian Matrix 20

3.6 The Fiedler Vector of a Graph 22

3.7 Extra Notes: Incidence Matrix of a Graph 23

3.8 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 Review For Quiz #1: Graphs and Spectral Decomposition **44**

7 Diagonalizable Matrices **45**

7.1 Similar Matrices 45

7.2 Eigenvectors as a Basis 46

7.3 Eigenvector Basis 47

7.4 Orthogonal Matrices 47

7.5 Symmetric Matrices 49

7.6 Skew-Symmetric Matrices 50

7.7 Nondiagonalizable Matrices 51

7.8 Extra Notes: Matrix Powers 53

7.9 Extra Notes: Small Perturbations 54

8 Spectral Decomposition and Positive [Semi–]Definite Matrices **56**

8.1 The Spectral Decomposition 56

8.2 Positive-Definite and Positive-Semidefinite Matrices 58

8.3 Quadratic Form of a Symmetric Matrix 58

8.4 Example: Product of a Full-Rank Matrix and its Transpose 59

8.5 Statistics of Vectors: Means and Variance 60

8.6 Example: Covariance Matrix in Statistics 61

8.7 Example: Linear Elastic Structures 63

9 Review For Test #1: Basic Linear Analysis **65**

10 Design Matrix and Standardized Data **67**

10.1 Data Matrix and Design Matrix 69

10.2 Standardized Data or Z Score 70

11 Orthogonal Projection **73**

11.1 Projecting to a 1D Subspace: Vector to Vector 73

11.2 Projecting to a 2D Subspace: Vector to Basis 75

12 Patterns – Linear Regression **78**

12.1 Residual Error 79

12.2 Linear Regression 80

12.3 Example: Hooke's Law 80

12.4 Linear Regression – One Dependent Variable, Plus Intercept 81

12.5 Data Matrix For Linear Regression 82

12.6 Assessment Of Linear Regression 84

12.7 Extra Notes For Derivations of Linear Regression 85

13 Cross-Validating Linear Regression **88**

13.1 Training and Testing 88

13.2 Linear Regression and the Design Matrix $\boldsymbol{X}$ 89

13.3 Measuring Error in Linear Regression 89

13.4 Cross-Validation of Linear Regression: Leaving Data Out 90

13.5 Cross-Validation of Linear Regression: K-Fold Analysis 91

13.6 Cross-Validation of Linear Regression: Monte Carlo Methods 92

13.7 Example: 13 Data With 2 Outliers 92

14 SVD – Singular Value Decomposition **95**

14.1 Eigenvectors of the matrix $[A^T A]$ 95

14.2 Eigenvectors of the matrix $[AA^T]$ 96

14.3 The Singular Value Decomposition, or SVD 97

14.4 Using the SVD 98

15 Review for Quiz #2: Matrices and Linear Regression **99**

16 Orthonormal Basis Vectors and the SVD **100**

16.1 SVD of a Square Matrix 100

16.2 SVD of a Non-Square Matrix 102

16.3 The SVD as an Approximate Basis for a Vector Space 103

16.4 Some SVD Properties 104

17 Principal Components Analysis – PCA **106**

17.1 Motivation, by Example 106

17.2 Zero-Mean Data Matrix 107

17.3 Principal Components Analysis as an SVD 108

17.4 Using the SVD to Compute PCA Scores 109

17.5 Matrix Norms 111

17.6 Eigenvalues and Singular Values 112

17.7 Extra Notes: A Matrix As A Series 113

17.8 Extra Notes: Eckart-Young Theorem 114

18 Review for Test #2: Matrices and Linear Regression **116**

19 PCA – Matrix Algebra and Dimensionality Reduction **118**

19.1 Scatter Matrix, SVD, and PCA 119

19.2 PCA and Low-Rank Approximation 120

19.3 PCA for Dimensionality Reduction 121

19.4 Approximations and The Scree Plot 121

20 Unsupervised Learning – K-Means Clustering **123**

20.1 Supervised and Unsupervised Classification 123

20.2 Unsupervised Classification: Data Clustering 123

20.3 Clustering – Iris Data 124

20.4 K-Means Clustering 125

20.4.1 A K-Means Algorithm 125

20.4.2 Example: Fisher's Iris Data 126

21 Classification – Linear Separability **130**
21.1 Separation of Two Clusters . . . 130
21.2 Separation of Three or More Clusters . . . 133
21.3 Hyperplane Specification With a Unit Normal Vector . . . 136

22 Classification – Assessment With Confusion Matrix **137**
22.1 Relative Confusion Matrix . . . 139

23 Classification – Assessment With ROC Curve **141**
23.1 Receiver Operator Characteristic – ROC . . . 141
23.2 Example: Two Variants of a Virus . . . 143

24 Review for Quiz #3: SVD, PCA, And Dimensionality Reduction **147**

25 Odds of Occurrence and Probability **148**
25.1 Logistic Function – Some Properties . . . 152
25.2 Hyperplane Classification and Pseudo-Distance . . . 153

26 Elementary Numerical Optimization **154**
26.1 Scalar Argument – Fixed-Stepsize Search . . . 155
26.2 Vector Argument – Fixed-Stepsize Search . . . 158

27 Review for Test #3: SVD, PCA, And Dimensionality Reduction **160**

28 Artificial Neuron – Learning Weights **162**
28.1 History – The Perceptron Rule . . . 163
28.2 Logistic Activation of an Artificial Neuron . . . 163
28.3 The Method of Steepest Descent . . . 164
28.4 Extra Notes: Gradient of Squared-Error Objective . . . 166

29 Classification – Logistic Regression **168**
29.1 Semilinear Activation – Logistic Function As A Sigmoid Curve . . . 169
29.2 Models Of Residual Error . . . 170
29.3 Implementations Of Logistic Activation . . . 172
29.3.1 Example – Fisher's Iris Data . . . 172

30 Nonlinear Separation – Embeddings and Gram Matrix **175**
30.1 Linear Separation Using an Embedding 176
30.2 Kernel Functions and the Gram Matrix 178
30.3 Some Kernel Functions for Row Spaces 179
30.4 Extra Notes for The Gram Matrix and Kernel Functions 181

31 Nonlinear Separation – Kernel PCA **182**
31.1 Principal Components Analysis And Scatter Of Observations 182
31.2 Kernel Functions and the Gram Matrix 185
31.3 Example – Kernel PCA For Fisher's Iris Data 186

32 Spectral Clustering Of Data **188**
32.1 Multiple Clusters In Graphs 188
32.1.1 Example: $\mathbf{30}$ Vertices In $\mathbf{3}$ Clusters 189
32.2 Graph Representation Of Simple Data 190
32.3 Weighted Graphs 194
32.4 Extra Notes on The Laplacian Matrix 194

33 Preparation for Quiz #4: k-Means, Assessment, and Odds of Occurrence **197**

34 Course Summary – Linear Data Analysis **198**

35 Preparation for Test #4: k-Means, Assessment, and Odds of Occurrence **202**

References **204**