Data Science Syllabus

Sources:

[1] Lars Eldén. *Matrix Methods in Data Mining and Pattern Recognition*, SIAM, 2007. https://doi.org/10.1137/1.9780898718867

[2] Calafiore & El Ghaoui. *Optimization Models*, Cambridge University Press, 2014. https://doi.org/10.1017/CBO9781107279667

[3] Guenin, Könemann & Tunçel. *A Gentle Introduction to Optimization*, Cambridge University Press, 2014. https://doi.org/10.1017/CBO9781107282094

[4] Higham & Higham. *Deep Learning: An Introduction for Applied Mathematicians*, SIAM Review, 2019. https://doi.org/10.1137/18M1165748

[5] Gilbert Strang. *Linear Algebra and Learning from Data*, SIAM, 2019. https://epubs.siam.org/doi/book/10.1137/1.9780692196380

[6] Robert J. Vanderbei. *Linear Programming: Foundations and Extensions*, Springer, 2020. https://link.springer.com/book/10.1007/978-3-030-39415-8

Books contain lots of good questions that can be used to prepare and to create exams. Moreover they have electronic versions available for free through UC Davis (and most universities).

1. Applied Linear Algebra (at the level of MAT 167)
1.1 Vectors, Matrices, Eigenvalues [1, Ch 2,15] or [2, Ch 1,2,3,7] or [5, Part I]
1.2 Singular Value Decomposition [1, Ch 6] or [2, Ch 5] or [5, Part I]
1.3 Linear Equations, Least Squares, and LU Decomposition [1, Ch 3] or [2, Ch 6,7] or [5, Part I]
1.4 Orthogonality, QR Decomposition, and Gram-Schmidt Projections [1, Ch 4,5] or [5, Part I,II]
1.5 Symmetric Matrices, PSD [1, Ch 3.2] or [2, Ch 4] or [5, Part I,II]

2. Optimization (At the level of MAT 168/170)
2.1 Convex Analysis, Gradient Descent [2, Ch 8] or [5, Part VI] or [3, Ch 7.3,7.4] or [6, Ch 10]
2.2 Linear Programming and its Duality [3, Ch 2,3] or [2, Ch 9] or [6, Ch 2,3]
2.3 Network Flows, Max-Flow Min-Cut Theorem [3, Ch 1.4,1.5,5.3] or [6, Ch 15]
2.4 General Optimality Conditions (Karush-Kuhn-Tucker) in Nonlinear Optimization [3, Sec 7.5,7.7]

3. Data/Machine Learning (at the level of MAT 170)
3.1 Basic Models for Supervised Learning and Regression [2, Sec 13.1,13.2] or [6, Ch 12]
3.2 Binary Classification Models (Support Vector Machines, Logistic Regression) [2, Sec 13.3]
3.3 Unsupervised Learning [2, Sec 13.5]
3.4 *k*-Means, Spectral Clustering [1, Ch 9] and [5, Part IV.7]
3.5 Basics of Neural Networks [4] or [5, Part VII]

Students are also expected to have some “hands-on” experience with programming and feel comfortable working with computers and data. (However, the prelim exam will not include programming.)