

2022 MIDAS Data Science for Biomedical Scientists Bootcamp (Dinov)

Time	Day1 Monday, July 18, 2022, 8:45AM-12:00PM US-EDT	Day 2 Tuesday, July 19, 2022, , 8:30-10:30 AM US-
7:00-8:30	<p>Session 1: Welcome and introduction to the program (Kayvan Najarian)</p>	<ul style="list-style-type: none"> • Review of Day 1, Questions, Comments, issues? • Estimate the square root function $\sqrt{\cdot}$ using NN • NN Google Trends and the Stock Market
8:45-9:00	<p>Session 2: Math foundations I – Brief introduction to mathematical foundations of machine learning</p> <ul style="list-style-type: none"> • SOCR Resources: Datasets & Case-studies, Webapps, DSPA, Spacekime/TCIU, GitHub, Prob & Stats EBook, SMHS EBook, Current SOCR Users • Open Science – <i>It's online, therefore it exists!</i> • Download DSPA Textbook (free) 	<ul style="list-style-type: none"> • Silhouette plots • Pediatric trauma clustering study • Text modeling & NLP (sentiment analysis example) • Longitudinal data analysis (Google trends analytics) • Deep neural networks (image-classification example)
9:00-9:30	<ul style="list-style-type: none"> • Math Foundations <ul style="list-style-type: none"> ○ Coordinate Systems (Cartesian, Polar, Spherical) ○ Calculus of Differentiation & Integration ○ Linear Algebra and Matrix Computing 	<p>Session 6: Linear regression, logistic regression</p> <ul style="list-style-type: none"> • Simple linear regression, multivariate LM, logit modeling • Ordinary least squares estimation • Regularized LM (LASSO, Ridge, Elastic Net) • Controlled feature selection (knockoff)
9:30-10:00	<ul style="list-style-type: none"> ○ Role of optimization in AI/ML ○ Optimization theory (Healthcare Manufacturer Product Optimization) 	
10:00-10:30	<ul style="list-style-type: none"> • Differential Equations 	
10:30-11:00	<ul style="list-style-type: none"> • Dimensionality reduction <ul style="list-style-type: none"> ○ Linear PCA: 2D→1D example, PPMI example ○ Non-linear: MNIST/OCR: UMAP OCR, t-SNE OCR ○ SOCR/Tensorboard/Projector UKBB Brain Study 	<p>Other Notes</p> <ul style="list-style-type: none"> • Motivation – the 7D of Big Data <ul style="list-style-type: none"> ○ Digitalization of all human experiences ○ Responsible Data Science/Ethical Predictive Analytics • R vs. Python vs. SAS vs. SPSS vs. other SW <p>Rmarkdown Notebook (IDE) – End-to-end Pipeline Workflow from raw data → models → visualization → analytics → reporting/pubs</p>
11:00-11:30	<ul style="list-style-type: none"> • EDA <ul style="list-style-type: none"> ○ SOCRAT (NI data of AD/MCI/NC), ○ Motion Charts (Housing Prices), • BrainViewer (raw MRI, DTI tracks, Brain Atlas)Probability Distributions – Distributome, TVN Webapp • Probability and Statistics 	
11:30-12:00	<ul style="list-style-type: none"> • Open discussion • Capstone project – interactive-learning using a large Autism data tensor ($n = 1,098$; $k = 2,145$). Use the RMD source, the example HTML output, and the provided data to experiment with some of the DSPA techniques. Think of ways to <i>augment</i> these data (e.g., expand the time range and increase the feature richness) 	<p>DSPA Appendices: Bayesian Simulation, Modeling and Inference » Information-Theoretic Foundation of Statistical Learning » Surface, Shape, and Manifold Representation and Visualization » Power Analysis in Experimental Design » Database SQL/NoSQL Queries & Google BigQuery » Image Convolution, Filtering, & Fourier Transform » Causality, Transfer Entropy, & Mechanistic Effects » Agent-based Reinforcement Learning</p>