SB220/SB221 Syllabus

SB220

12 classes

Week	Date	Module	Topic	Related P-Set	TF Section	Reading
1	1/29		Introduction, motivation			
	1/31	Estimation	Quantitative estimation: how to lie skillfully	0.1	2/3: Quantitative estimation examples	[1] Biology by the Numbers, Phillips & Milo
2	2/5	Stats	Intro to statistics: concepts surrounding measurement error; metrics of assay performance; estimators; bias and error, jackknife and bootstrap procedures	1.1		[2] An Introduction to Statistical Learning, James, Witten, Hastie, Tibshirani. http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf Chapters 1,2

	2/7		Intro to statistics (cont'd from prev. class). The physical origin of key distributions seen in biology: Poisson statistics and shot noise; exponential distributions; binomials; negative binomials; the Central Limit Theorem		2/10: Hand in P-Set 0; Bootstrap implementation	
3	2/12		The physical origin of key distributions seen in biology (cont'd from prev. class).	1.2		TBD
	2/14		Error propagation 1: Numerics and analytical warm-up	1.3	2/17: President's Day. No section.	TBD
4	2/19		Error propagation 2: Theory of error propagation; analytical approaches; pitfalls of correlated errors; further numerical approaches			

	2/21		Hypothesis testing: type I and II errors; back-of-the-envelope T-tests; build-your-own non-parametric tests	2	2/24: Hand in P-Set 1	TBD
5	2/26		Multiple hypothesis testing: the problem; error rates; Bonferroni and Benjamini-Hochberg corrections	2		Class notes
	2/28	Applications	P-hacking. Example on RNA-Seq data		3/2: Hand in P-Set 2	TBD p-hacking perspective/review
6	3/4		Worked out example: Single molecule FISH - error propagation, distributions, sources of error			
	3/6		Guest lecture on cryoEM		3/9: More worked out examples	
	3/11		NO CLASS			

SB221

12 classes

Week	Date	Module	Topic	Related P-Set	TF Section	Reading
7	3/13	Statistics in High Dimensions	Motivation: high-dimensional biological data; Properties of high-dimensional spaces; Normal distributions in high-dimensional space;Basics of non-euclidean geometry		3/23: High-dimensional data in whole genome sequencing; single cell RNA-Seq; multi-color FACS; metabolomics; multiplexed imaging modalities; Mass Spectrometry; Neural activity arrays; patient records	TBD
Spring Break (3/16 - 3/20)						
8	3/25	Statistics in High Dimensions	Linear Algebra review; eigenvectors and eigenvalues; rank; calculus on matrices			
	3/27		Linear dimensionality reduction; PCA, SVD, Kernel PCA		3/30: Analysis of a biological data set	

9	4/1		Nonlinear dimensionality reduction: MDS, Isomap, Spectral methods			
	4/3		Nonlinear dimensionality reduction: tSNE, UMAP		4/6: Computational exploration of methods on a biological dataset	Problem Set
10	4/8	Statistical Learning	Introduction to statistical learning; Bayesian inference; classifiers LDA;			David McKay's textbook
	4/10		Logistic regression; model selection; bias versus variance; machine learning intro		4/13: Bayesian thinking (e.g disease screening); Implement your own classifier	
11	4/15		Neural networks, back propagation; practical considerations for using neural networks; cross-validation; p-hacking in ML			
	4/17		Application lecture: CNNs, image processing, denoising.	3	4/20: Worked-out Python notebook implementing TensorFlow	

12	4/22		Supervised PCA; auto-encoders for dimensionality reduction; VAE and GANs			
	4/24	Biophysics of measurements	Physics of inference. Can biological systems do bayesian inference? reversible and irreversible transitions; equilibrium vs energy dissipation; kinetic proofreading		4/27:Hand in P-Set 3	
13	4/29		Examples of biological systems carrying out measurement: rhodopsin; nucleotide detection by DNA Polymerase; T-cell receptors; GPCRs detecting ligand.			