Course Rubric & Title:
PHYS 535: Physics-inspired Statistical Data Analysis and Stochastic Processes in Physics

Course Description:
This course covers modern physics-inspired mathematical, statistical and Monte Carlo methods for analyzing scientific data. Topics to be covered include review of linear algebra and Hilbert space, spectral graph theory, clustering methods, dimensional reduction techniques, Reproducing Kernel Hilbert Space, kernel embedding, Grassmannian manifolds, matrix and tensor decompositions, stochastic sampling methods, numerical optimization, cross entropy method, Markov Chain Monte Carlo, and Gaussian Process.

Prerequisites and Corequisites:
Strong background in linear algebra, analysis, statistical mechanics, classical mechanics, and quantum mechanics

Credit Hours: 4

Course Instructor:
Jun S. Song

Course Website (for registered students only):
https://canvas.illinois.edu

Texts and Supplies:
Required: Lectures notes prepared by the instructor.

Recommended: *The Elements of Statistical Learning*, by Hastie, Tibshirani, and Friedman.
Machine Learning: A Probabilistic Perspective, by Kevin Murphy
Matrix Computations, by Golub and Van Loan.

Academic Integrity

All activities in this course are subject to the Academic Integrity rules as described in Article 1, Part 4, Academic Integrity, of the Student Code.

Infractions include, but are not limited to:

- Cheating, plagiarism, fabrication
- facilitating infractions of academic integrity
- academic interference
- computer-related infractions
- unauthorized use of university resources
- sale of class materials or notes

Violations of any of these rules will be prosecuted and reported to the student's home college in compliance with the Student Code: Article 1, Part 4, Academic Integrity, of the Student Code.

All aspects of the course are covered by these rules, including:

- homework

Course Grading:

Course grading will proceed in compliance with University policy as given in Article 3, Part 1 of the Student Code.

Gradebook

Students will be able to view their grades on all components of the course using the course grade book on Canvas.

Students are responsible for reporting any discrepancies found in their student grade book to the attention of their section instructor immediately.

Grading

60% Homework, 40% Term Project

Final Grade

Final course grade will follow a Bell curve, with the mean centered between B+ and A-.

Tardiness

Late homework will not be accepted, unless there is a medical reason.

Course Completion

In compliance with the Student Code, the instructor of this course determines the amount of coursework which must be completed to pass.

Disability Access
(https://www.disability.illinois.edu/academic-support/instructor-information/examples-disability-statements-syllabus)

The Department of Physics is committed to being an open and welcoming environment for all of our students. We are committed to helping all of our students succeed in our courses.

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak St., Champaign, call 333-4603, e-mail disability@illinois.edu or go to the DRES website. If you are concerned you have a disability-related condition that is impacting your academic progress, there are academic screening appointments available on campus that can help diagnosis a previously undiagnosed disability by visiting the DRES website and selecting “Sign-Up for an Academic Screening” at the bottom of the page.

Course Component Breakdown:

Lectures: 2 per week for 15 weeks (80 minute lectures)

Course Topics and Learning Objectives:

Lecture 1: Introduction to the course	
	Discuss problems associated with high-dimensional data. Define dimensional reduction, embedding, stochastic sampling and optimization. Introduce the concepts behind feature maps and support vector machine.
Lecture 2: Fundamentals of Linear Algebra	
	Define spanning set, basis, function space, subspace, inner product, norm, metric, linear maps, kernel, image, and co-kernel. Derive matrix representation of linear maps. Prove the Riesz representation theorem in finite dimensions.
Lecture 3: Eigenvalues and Singular Values	
	Understand eigenvalues and eigenvectors. Define positive definite functions and kernels. Understand power method for approximating dominant eigenvalues and eigenvectors. Derive singular value decomposition.
Lecture 4: Operator Norms and Matrix Approximations	
	Define vector space of matrices. Understand low-rank matrix approximation. Define operator norm and bounded operators.
Lecture 5: Dimensional Reduction Techniques (PCA)	
	Derive Principal Component Analysis (PCA). Define Rayleight quotient and understand its connection to PCA Understand generalized eigenvalue problems.
Lecture 6: Dimensional Reduction Techniques (Multidimensional Scaling)	
	Define isometric embedding. Understand exact and approximate multidimensional scaling.
Lecture 7: Dimensional Reduction Techniques (Spectral Embedding)	
	Define graph Laplacians and understand connections to classical mechanics. Understand spectral embedding.
Lecture 8: Dimensional Reduction Techniques (Stochastic Neighbor Embedding)	
	Understand t-SNE and the associated optimization algorithms.
Lecture 9: Reproducing Kernel Hilbert Space (Finite Dimensional)	
	Understand finite dimensional Reproducing Kernel Hilbert Space (RKHS). Learn how to obtain feature maps from reproducing kernels.
Lecture 10: Reproducing Kernel Hilbert Space (Infinite Dimensional)	
	Define Cauchy sequences and Hilbert space. Define separable Hilbert space. Define infinite dimensional RKHS. Discuss Riesz Representation Theorem and Moore-Aronszajn Theorem.
Lecture 11: Reproducing Kernel Hilbert Space (Kernel Construction)	
	Define Mercer Kernels. Understand connections to Green's functions in physics.
Lecture 12: Reproducing Kernel Hilbert Space (Applications)	
	Understand ridge regression and kernel regression. Understand kernel support vector machine.
Lecture 13: Multi-layer Graphs (Introduction)	
	Define graph kernel. Define Stiefel and Grassmannian manifolds. Define principal angles.

Lecture 14: Multi-layer Graphs (Algorithm)
Understand the computation of principal angles via optimization. Understand the computation of principal angles via SVD. Define distance between subspaces.
Lecture 15: Multi-layer Graphs (Applications)
Understand multi-layer spectral clustering. Understand Pareto multi-objective optimization.
Lecture 16: Tensor Decompositions (Introduction)
Define free vector space and tensor product space. Define transformation of tensors. Define multilinear maps.
Lecture 17: Tensor Decompositions (Tensor Rank and Tensor Unfolding)
Define tensor rank and CP decomposition of tensors. Define symmetric and alternating tensors. Understand tensor unfolding.
Lecture 18: Tensor Decompositions (Tensor Embedding and Approximation)
Understand tensor representation of data. Understand iterative CP approximation algorithm.
Lecture 19: Tensor Decompositions (Higher-Order SVD and Matrix Product States)
Understand Higher-Order SVD. Understand matrix product states.
Lecture 20: Tensor Decompositions (Tensor Networks and Tensor Eigenvalues)
Define tensor networks. Understand eigenvalues and eigenvectors of tensors.
Lecture 21: Monte Carlo Sampling Techniques (Introduction)
Explain the motivation. Understand transformation of random variables. Discuss direct inversion of cumulative distributions.
Lecture 22: Monte Carlo Sampling Techniques (Rejection Sampling)
Define rejecton sampling. Understand the efficiency of rejection sampling. Understand Bayesian posterior sampling. Define nonhomogeneous Poisson processes.
Lecture 23: Monte Carlo Sampling Techniques (Importance Sampling)
Define biased and unbiased estimators. Introduce cross-entropy and KL divergence. Understand importance sampling.
Lecture 24: Monte Carlo Sampling Techniques (Markov Chain Monte Carlo)
Define Markov Chain. Define MCMC. Understand simulated annealing and parallel tempering.
Lecture 25: Monte Carlo Sampling Techniques (Markov Chain Monte Carlo)
Understand Metropolis-Hastings algorithms. Understand Gibbs sampling.
Lecture 26: Numerical Optimization (Introduction)
Define combinatorial optimization. Define convex optimization. Introduce genetic algorithms.
Lecture 27: Numerical Optimization (Cross Entropy Method)
Define rare event sampling.

	Understand cross entropy method and its connection to importance sampling.
Lecture 28: Numerical Optimization (Applications)	
	Learn applications of the cross entropy method to combinatorial optimization problems. Understand the technique of Hamiltonian Monte Carlo.
Lecture 29: Gaussian Process	
	Define Schur complement. Understand mariginal and conditional distributions of multivariate normal distributions. Understand Gaussian Process regression.