

Georgia State University

Department of Mathematics and Statistics

Topics of Statistical Learning

Course add proposal and sample syllabus

Prepared by Chi-Kuang Yeh | 21 October 2026

Proposed graduate elective for advanced master's and doctoral students. The course develops statistical methods for learning with incomplete, noisy, comparative, or sequential information and for inference after data-driven selection.

Course information

Subject and number: STAT 8XXX; number to be assigned.

Long title: Topics of Statistical Learning

Short title: Topics Statistical Learning

Credits and format: 3 credit hours; lecture and seminar with integrated R laboratories.

Grade mode: Proposed letter grading; authorized administrative code to be assigned.

Curricular role: Graduate elective in statistics, statistical learning, and data science.

Course fee: None proposed. Required activities use ordinary laptop computing.

Catalogue description

Selected topics in statistical learning with incomplete, noisy, comparative, and sequential feedback. Topics include positive-unlabelled and unlabelled-unlabelled learning, other forms of weak supervision, preference learning, online learning and bandits, and selective inference after model selection and clustering. Emphasis is placed on observation mechanisms, identifiability, risk estimation, computational methods, theoretical guarantees, and reproducible empirical evaluation. Students complete mathematical and computational assignments, discuss research papers, and carry out a focused methodological project. Specific applications and advanced topics may vary by offering.

Rationale

Many learning problems provide only part of the information assumed in standard supervised learning: a few confirmed positives, uncertain labels, comparisons rather than absolute responses, or feedback acquired through sequential decisions. A second challenge arises when the same observations are used to discover a structure and to assess its statistical significance. This course connects these problems through explicit targets, observation models, assumptions, and guarantees. It prepares students to read methodological papers and evaluate when a learning or inferential procedure supports the scientific claim being made.

Curriculum and delivery

Preparation and intended audience

Proposed prerequisites are graduate standing and preparation equivalent to mathematical statistics, regression or generalized linear models, linear algebra, multivariable calculus, and basic R programming. Students should be able to work with conditional expectation, likelihood, elementary asymptotics, and numerical optimization. A first course in statistical or machine learning is recommended. Equivalent preparation may be established through instructor permission; the department should translate these competencies into registration rules.

The first offering is designed for advanced master's and doctoral students. A fully developed undergraduate version would require a separate reduction in proof depth and reading load. Measure-theoretic probability, reinforcement learning, and large-model training are not prerequisites.

Relationship to other courses

An introductory statistical learning course develops a broad set of prediction methods. This course builds on that preparation by studying how the form and collection of supervision affect identifiability, estimation, sequential performance, and inferential validity. A computational statistics course supplies simulation and optimization tools that are used here to investigate methodological claims. An applied capstone develops the complete analysis workflow; this course instead organizes projects around a focused statistical question about a method or assumption.

The department should compare the proposal with its current statistical learning, high-dimensional statistics, computational statistics, nonparametric methods, and inference offerings. These are intended curricular distinctions; a complete review of present course overlap is part of departmental consultation.

Teaching format and resources

Proposed delivery is face-to-face, with two meetings per teaching week combining lectures, derivations, guided R exercises, and paper discussions. A typical week has one central statistical question, one selected reading or instructor note, and one short computational illustration. Readings are assigned in excerpts; students are not expected to read complete textbooks during the term.

R is the primary teaching language. Quarto or R Markdown supports reproducible submissions. Required experiments use small simulated datasets, finite-dimensional models, and documented author software. An instructor-approved alternative language may be used for a project when it is reproducible and does not require additional course support. No paid API, AI subscription, or GPU service is required.

Scheduling and consultation

Proposed frequency is once every two years, subject to demand and staffing. No programme requirement or repeat-credit entitlement is assumed. Consultation should include graduate programme representatives and instructors of related learning and inference courses. Consultation has not yet been recorded as completed.

Administrative completion

Before formal submission or student distribution, assign the course number, first term, administrative codes, and meeting details; confirm prerequisites and permitted graduate grade symbols; document consultation; and insert the current institution-required syllabus statements. The academic plan below is specified independently of those term-specific fields. GSU's official syllabus guidance provides the completion requirements.

[GSU syllabus guidance](#)

Sample syllabus

STAT 8XXX | Topics of Statistical Learning | 3 credit hours

Instructor: Chi-Kuang Yeh | Email: cyeh@gsu.edu

Term and year: [To be assigned] | CRN: [To be assigned]

Meetings and location: [To be assigned]

Office and office hours: [To be completed before registration]

Course objective

Students will learn to turn a nonstandard learning setting into a statistical problem with an explicit target, an observation mechanism, and a defensible assessment strategy. The course combines derivations, simulation, implementation, and critical reading. Weak supervision, sequential learning, and selective inference are related through these questions, while retaining their different mathematical objectives.

Learning outcomes

1. Specify the observed data, latent quantities, sampling unit, and target risk or inferential target in a nonstandard learning problem.
2. Derive PU and two-mixture UU risk identities and identify the assumptions required to use them.
3. Distinguish missing labels, corrupted labels, candidate labels, complementary labels, and comparative preferences.
4. Fit and assess a pairwise preference model, including its identification constraints and sources of dependence.
5. Define an appropriate regret criterion, derive one online gradient descent bound, and compare sequential learning algorithms.
6. Explain why data-driven selection affects inference and implement a justified test after a specified clustering procedure.
7. Design a reproducible methodological study, evaluate uncertainty and assumption sensitivity, and communicate the scope of its conclusions.

Materials and software

Instructor notes, selected articles, R examples, datasets, and assignment rubrics will be posted through iCollege. The numbered readings at the end of this proposal identify sources for the weekly plan. No single textbook covers the full course, and no textbook purchase is required.

Students should be comfortable writing functions, simulating data, fitting regression models, and using an optimizer in R. Each submission should record its random seeds, software versions, data-generation or preprocessing steps, and execution instructions. Public repositories are optional; reproducible files may be submitted through the course system.

Workload and scope

Plan for approximately six to nine hours of preparation and independent work per week, with variation around project deadlines. The common theory core is deliberately limited: risk rewriting for PU and UU learning, a preference likelihood, one regret proof, and one elementary selection-adjusted pivot. More advanced results are interpreted and explored rather than all proved from first principles.

Weekly plan for weeks 1 to 8

The schedule contains 15 teaching weeks. Calendar dates, holidays, and the final assessment period will be set for the term of offering. Numbers in brackets refer to the reading list. Core readings mean selected sections, not an entire book or paper.

Wk	Focus	Statistical and computational work	Reading or milestone
1	Statistical learning foundations	Target risk, empirical risk minimization, identifiability, generalization, and observation mechanisms. Distinguish prediction error from inferential error.	Instructor review notes; readiness exercise.
2	Forms of weak supervision	Noisy labels as a worked example; transition mechanisms. Overview of partial, complementary, bag-level, and multiple-source labels.	[1]; optional [13], [14].
3	Positive unlabelled learning	Observed versus latent labels; representative positives and marginal unlabelled data; SCAR; derive the PU risk identity.	[2], [3]; Assignment 1 released.
4	PU estimation and sensitivity	Unbiased and non-negative empirical risks; class-prior misspecification; feature-dependent positive selection.	[3]; selected [12].
5	Unlabelled unlabelled learning	Two mixtures with distinct known class priors and shared class-conditionals; target prior; inversion and instability.	[4]; Assignment 1 due.
6	Preference models	Bradley-Terry likelihood, comparison graphs, identifiability, finite estimates, and interpretable contrast estimation.	[5]; project proposal due.
7	Preference data in practice	Context effects, order bias, ties, repeated prompts and annotators; pairwise prediction versus ranking.	[5]; optional [15]; Assignment 2 released.
8	Full information online learning	Sequential protocol, expert aggregation, Hedge and projected gradient descent; derive one fixed-comparator regret bound.	[6], selected introductory and OGD sections.

How the first two blocks connect

Weeks 2 to 5 ask how a risk can be recovered from incomplete or corrupted supervision. Weeks 6 and 7 replace absolute labels with comparisons. Week 8 changes the timing of decisions and feedback. A comparison dataset can be collected in a fixed batch; preference learning does not itself imply an online protocol.

Weekly plan for weeks 9 to 15

Wk	Focus	Statistical and computational work	Reading or milestone
9	Bandit feedback	Finite stationary bandits, exploration, UCB, cumulative pseudo-regret, and best-arm identification as a different objective.	[7], Chapters 4, 6 and 7.
10	Duelling bandits	Adaptive comparison selection; specify a winner criterion and sequential objective; connect comparison allocation to information.	[8]; Assignment 2 due.
11	Inference after selection	Data-dependent targets and hypotheses; conditional error control; splitting when justified; a truncated Gaussian pivot.	[9]; project reproduction checkpoint due.
12	Hierarchical clustering	Selected cluster mean contrasts; known-variance Gaussian model; naive versus selective p-values.	[10]; author tutorial; Assignment 3 released.
13	K means and validation	Algorithm-path conditioning; appropriate independent validation; distinguish cluster significance, stability, and recovery.	[11]; optional [16].
14	Synthesis and project clinic	Compare assumptions and guarantees across modules; revise project claims. At most one short optional extension if time permits.	Assignment 3 due; project draft discussion.
15	Project presentations	Present a focused result, defend the evaluation design, and identify limits of the method or empirical evidence.	Presentations; final report in the designated final period.

Depth for the first offering

The course requires students to derive one elementary selective Gaussian pivot and to interpret the clustering methods' assumptions. Full derivations of every hierarchical-linkage or k-means truncation constraint are beyond the common core. Similarly, UCB and duelling-bandit guarantees are introduced through their assumptions, objectives, and empirical behaviour; a complete reinforcement learning course is not folded into this block.

Rotating extensions

Possible extensions include weak-label aggregation, class-prior estimation, active learning, contextual or nonstationary bandits, preference-based language-model training, data thinning, and conformal prediction. An offering may select one or two for readings or projects. Conformal predictive coverage and selective Type I error control remain distinct guarantees. Week 14 can be reserved entirely for project work when needed.

Assessment and course policies

Assessment	Weight	Timing
Three integrative assignments	45%	15% each; weeks 5, 10 and 14
One paper discussion and reading note	10%	Assigned during weeks 6 to 13
Project proposal	5%	Week 6
Project reproduction checkpoint	10%	Week 11
Final report and reproducible package	20%	Designated final assessment period
Project presentation and individual questions	10%	Week 15
Total	100%	No written midterm or final examination

Proposed grading scale: A = 90 to 100; B = 80 to below 90; C = 70 to below 80; D = 60 to below 70; F = below 60. The department will confirm permitted graduate grade symbols and any plus/minus subdivision before the syllabus is issued.

Individual work and participation

Assignments are individual. Discussion of concepts is encouraged, but submitted derivations, code, and interpretation must reflect the student's own work and acknowledge substantive collaboration. Students are expected to participate in lectures, laboratories, and peer discussion. Attendance has no separate percentage allocation; the scheduled paper discussion and project presentation are assessed activities.

Extensions and missed assessments

Students should request an extension before the deadline when possible. Approved accommodations, emergencies, and excused absences will be handled individually, with a documented revised deadline or equivalent assessment. Unapproved late work is subject to a proposed deduction of 10 percentage points for up to 48 hours; later submissions require an approved completion plan. There are no written examinations to make up.

Academic honesty and AI assistance

Submitted work must accurately identify sources, authorship, reused code, and permitted assistance. AI tools may assist with debugging and language editing unless an assignment states otherwise. Students must verify all calculations, references, and interpretations and briefly disclose substantive assistance. Designated independent derivations and oral explanations must demonstrate the student's own understanding. University academic-honesty requirements apply.

Access and course administration

Students seeking accommodations should contact the university's Access and Accommodation Center and provide the approved plan. Course logistics, due dates, feedback, and authorized changes will be communicated through iCollege. Current institutional wording for accommodations, academic honesty, course evaluation, attendance reporting, and other required statements must be inserted before student distribution.

The course syllabus provides a general plan for the course; deviations may be necessary.

Assignment briefs

Assignment 1 Learning without complete labels

Students derive the PU risk identity under an explicitly stated sampling scheme, then implement a simple classifier in R. Compare treating unlabelled cases as negative, an unbiased PU objective, and a non-negative PU objective. A fully labelled classifier may be included as a clearly identified oracle benchmark. Vary a supplied class prior or the positive-selection mechanism, and evaluate on an independent labelled test sample reserved for assessment. [2, 3]

A short UU component derives the two-mixture inversion. For a fixed classifier, hold the shared class-conditional distributions constant and vary the separation between the two known mixture priors. Plot the variability of the estimated target risk. The goal is to connect identification with finite-sample stability, without adding a second extensive software project. [4]

Submission: one derivation, a reproducible R report, two or three focused figures, and an explanation of an assumption failure. Suggested length is 5 to 7 pages excluding code and references.

Assignment 2 Comparisons and sequential decisions

Fit a Bradley-Terry model to simulated pairwise outcomes under connected, sparse, and disconnected comparison designs. State an identification constraint and explain which contrasts are estimable. Include a separate example in which separation prevents a finite unpenalized fit despite graph connectivity. Use independent comparisons in the core exercise. [5]

Implement uniform exploration and UCB for a small stationary Bernoulli bandit. Report cumulative pseudo-regret, arm allocations, and Monte Carlo uncertainty. The simulator may use true arm means to evaluate performance; the learning algorithms may not. A short written question explains how choosing pairs instead of arms changes the feedback and objective in a duelling problem. [7, 8]

Submission: the likelihood and graph discussion, one bandit comparison, and a reproducible report of 5 to 7 pages. A provided simulation harness keeps the programming burden manageable.

Assignment 3 Inference after clustering

Generate a single Gaussian cloud with known noise scale, use a pre-specified two-cluster hierarchical procedure, and compare naive and selective p-values over repeated simulations. Fix preprocessing, linkage, cluster number, and the tested-pair rule in advance. Use the authors' documented R implementation for the selective calculation. [10]

Report null rejection rates with Monte Carlo uncertainty, p-value plots, and the precise selected mean-contrast null. Explain why a stable partition or an ordinary test on the same clustering features does not establish valid inference. A k-means comparison is optional and must preserve the algorithmic information required by its method. [11]

Paper discussion

Each student leads one short discussion and submits a two-page note stating the target, observed data, essential assumptions, central result, and one useful question or failure case. The discussion is assessed on understanding and explanation, rather than the number of slides or breadth of literature surveyed.

Research project

The project is a focused methodological investigation. Students may extend one assignment experiment, identifying what is reused and what is new. A successful project reproduces a defensible result and investigates one consequential assumption, design choice, or extension. A novel theorem or publishable method is not required.

Possible project directions

Direction	A manageable research question	Expected evidence
PU or UU learning	How do prior errors, positive-selection bias, or nearly equal mixture priors affect learning?	Risk or prediction-error curves; uncertainty across replications; explicit identification assumptions.
Preference learning	How do comparison allocation, order effects, or annotator heterogeneity alter estimates?	Comparison design; estimable contrasts; predictive performance; justified uncertainty analysis.
Online learning	What changes when rewards or preferences are observed adaptively?	A defined regret or identification target; allocation curves; at least one controlled sensitivity study.
Selective inference	When are selected cluster mean tests calibrated, and what happens after an assumption violation?	A precise selected null; null rejection rates; selection rule; qualified power or recovery analysis.

Required project components

- A clear statistical question, sampling model, estimand or performance objective, and scope of the claim.
- A justified baseline and one comparison or extension, supported by at least three relevant methodological sources.
- A simulation or evaluation design that distinguishes fitting, tuning, and final assessment and reports appropriate uncertainty.
- One deliberate sensitivity study or failure case, with an explanation of why the result changes.
- Reproducible code, a concise execution guide, and an account of software, data provenance, and any substantive assistance.

Milestones and evaluation

The week 6 proposal is approximately two pages. The week 11 checkpoint includes a working baseline, one reproduced figure or table, and a revised analysis plan. The final report is 8 to 10 pages excluding references and code, accompanied by a 12-minute presentation and individual questions.

The final report rubric is: problem formulation and assumptions, 25%; methodological reasoning, 25%; empirical design and uncertainty, 30%; reproducibility and communication, 20%. These are internal weights within the report component. An approved pair project must specify individual contributions, and each student answers questions individually.

Instructor notes on statistical scope

PU and UU identification

The standard two-sample PU derivation uses representative positives and an unlabelled sample from the full feature marginal. In a single partially labelled dataset, the subset with no observed label is generally not that marginal. Teach the sampling scheme before writing the empirical objective. SCAR does not by itself identify an unknown class prior or labelling probability. Non-negative risk correction can introduce finite-sample bias and need not dominate every alternative in every example. [2, 3, 12]

The core UU model uses two mixtures with known, distinct positive-class proportions and common class-conditional distributions. The target-population prior must also be specified, or a balanced target risk must be explicitly chosen. Risk inversion becomes unstable as the mixture proportions approach each other. Arbitrary shifts in class-conditionals invalidate that argument. UU is supervised through aggregate information; its classification target is not automatically supplied by clustering. [4]

Other weak labels

A partial label is a candidate set, a complementary label excludes a class, and a bag label refers to a group. These observations require different mechanisms. Use at least three classes for a genuinely weak complementary-label example: with two classes, a correct complementary label reveals the class. For noisy labelling functions, source dependence and abstention matter; repeated agreement is not automatically independent evidence. [1, 13, 14]

Preference uncertainty and online guarantees

Distinguish identification of strength differences from existence of a finite unpenalized Bradley-Terry estimate. Repeated prompts or annotators can induce dependence. Resampling or hierarchical modelling must respect the independent units; crossed dependence may need more than a one-way bootstrap. Adaptive comparisons require inference justified for the sampling history and intended stopping rule. An ordinary independent-data Wald interval is not automatically adequate. [5]

Online learning specifies a sequential protocol and comparator. Bandit learning specifies restricted feedback, and duelling bandits use pairwise feedback. These categories overlap but are not synonyms for reinforcement learning. A fixed-comparator regret guarantee does not establish adaptation to arbitrary drift or validity under optional stopping. The winner criterion must be stated before evaluating a duelling-bandit method. [6-8]

Inference after clustering

Under the introductory independent Gaussian model with known noise covariance, the cluster test concerns equality of the average underlying means of the observations in two selected groups. It does not by itself determine the number of mixture components or establish biological subtypes. The selection event and any nuisance conditioning must match the implemented algorithm. [10, 11]

For the laboratory, pre-specify the cluster number, algorithm, initialization rule, feature transformation, and tested pair. Retain the intermediate assignments for the k-means procedure. Splitting rows and assigning held-out rows using the same features subsequently tested may still create selection. Explain the independence needed by a validation design. Cluster stability, predictive coverage, selected-pair error control, and simultaneous error control answer distinct questions. [9-11, 16]

Core reading list

Assign the indicated sections or excerpts alongside instructor notes. The list supplies the main methodological anchors; it is not a requirement to read every item in full.

- [1] Natarajan, N., Dhillon, I. S., Ravikumar, P. K., and Tewari, A. (2013). Learning with Noisy Labels. Advances in Neural Information Processing Systems 26. Observation model and loss correction. [Primary source](#)
- [2] Elkan, C., and Noto, K. (2008). Learning Classifiers from Only Positive and Unlabeled Data. KDD, 213-220. Positive selection and the SCAR relationship. [Primary source](#)
- [3] Kiryo, R., Niu, G., du Plessis, M. C., and Sugiyama, M. (2017). Positive-Unlabeled Learning with Non-Negative Risk Estimator. Advances in Neural Information Processing Systems 30. Risk rewriting and non-negative correction. [Primary source](#)
- [4] Lu, N., Niu, G., Menon, A. K., and Sugiyama, M. (2019). On the Minimal Supervision for Training Any Binary Classifier from Only Unlabeled Data. ICLR. Two-mixture identification and risk estimation. [Primary source](#)
- [5] Turner, H., and Firth, D. (2012). Bradley-Terry Models in R: The BradleyTerry2 Package. Journal of Statistical Software, 48(9), 1-21. Likelihood, covariates, and implementation. [Primary source](#)
- [6] Hazan, E. (2022). Introduction to Online Convex Optimization, second edition. MIT Press. Selected introductory sections and the online gradient descent chapter; an updated author manuscript is available. [Primary source](#)
- [7] Lattimore, T., and Szepesvari, C. (2020). Bandit Algorithms. Cambridge University Press. Selected material from Chapters 4, 6, and 7. [Primary source](#)
- [8] Yue, Y., Broder, J., Kleinberg, R., and Joachims, T. (2012). The K-armed dueling bandits problem. Journal of Computer and System Sciences, 78(5), 1538-1556. Pairwise sequential feedback and objectives. [Primary source](#)
- [9] Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. The Annals of Statistics, 44(3), 907-927. Selection conditioning and a Gaussian pivot. [Primary source](#)
- [10] Gao, L. L., Bien, J., and Witten, D. (2024). Selective Inference for Hierarchical Clustering. Journal of the American Statistical Association, 119(545), 332-342. Published online in 2022; volume year 2024. [Primary source](#)
- [11] Chen, Y. T., and Witten, D. M. (2023). Selective inference for k-means clustering. Journal of Machine Learning Research, 24(152), 1-41. Algorithm-specific selective testing. [Primary source](#)

Further readings and teaching resources

These readings support selected extensions or projects. They are not additional weekly requirements.

- [12] Bekker, J., Robberechts, P., and Davis, J. (2019). Beyond the Selected Completely At Random Assumption for Learning from Positive and Unlabeled Data. ECML-PKDD. Feature-dependent positive selection. [Primary source](#)
- [13] Ishida, T., Niu, G., Menon, A., and Sugiyama, M. (2019). Complementary-Label Learning for Arbitrary Losses and Models. ICML, PMLR 97, 2971-2980. [Primary source](#)
- [14] Ratner, A. J., De Sa, C. M., Wu, S., Selsam, D., and Re, C. (2016). Data Programming: Creating Large Training Sets, Quickly. Advances in Neural Information Processing Systems 29. [Primary source](#)
- [15] Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. (2023). Direct Preference Optimization: Your Language Model is Secretly a Reward Model. Advances in Neural Information Processing Systems 36. Use a finite-action example or supplied log probabilities. [Primary source](#)
- [16] Neufeld, A., Dharamshi, A., Gao, L. L., and Witten, D. (2024). Data Thinning for Convolution-Closed Distributions. Journal of Machine Learning Research, 25(57), 1-35. [Primary source](#)
- [17] du Plessis, M. C., Niu, G., and Sugiyama, M. (2015). Convex Formulation for Learning from Positive and Unlabeled Data. ICML, PMLR 37, 1386-1394. Supplementary instructor reference for PU risk estimation. [Primary source](#)
- [18] Bradley, R. A., and Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons. Biometrika, 39(3-4), 324-345. Historical and likelihood foundation. [Primary source](#)

R resources

Preference modelling: [BradleyTerry2 article and code](#)

Hierarchical clustering inference: [clusterpval documentation](#)

K-means inference: [KmeansInference documentation](#)

The authors' clustering software should be exercised in the teaching environment before release of the laboratory. Fix the package versions and provide a minimal worked example with the same settings used in the assessment.

Preparing the first offering

Prepare one reusable R experiment for each assessed block, one concise derivation note for each main theoretical anchor, and a common project rubric. Allow projects to extend assignment work. This keeps the course coherent while leaving room to update readings and applications in later offerings.