Boston University College and Graduate School of Arts & Sciences
Undergraduate Academic Program Office
725 Commonwealth Avenue, Room 102

CAS/GRS Course Revision Proposal Form

This form is to be used when proposing a revision of an existing CAS or GRS course.

Once completed, this form should be submitted to Senior Academic Administrator Peter Law (617-353-7243) as a PDF file to pgl@bu.edu.

For further information or assistance, contact Associate Dean Joseph Bizup (617-353-2409; jbizup@bu.edu) about CAS courses or Associate Dean Jeffrey Hughes (617-353-2690; hughes@bu.edu) about GRS courses.

DEPARTMENT OR PROGRAM: Mathematics and Statistics, MSSP

CURRENT COURSE NUMBER: GRS MA 685 B1
CURRENT COURSE NAME: Advanced Topics in Applied Statistical Analysis

CURRENT 40 WORD COURSE DESCRIPTION:
Continues topics of GRS MA 684 at a more advanced level. Canonical correlation, multivariate analysis of variance, multivariate regressions. Categorical dependent variables techniques; discriminant analysis, logistic regression, log-linear analysis. Factor analysis; principal-axes, rotations, factor scores. Cluster analysis. Power analysis. Extensive use of statistical software. This course cannot be taken for credit in addition to the course with the same title that was previously numbered CAS MA 685.

{Please Note: 685 is run for both MSSP-only and non-MSSP audiences; this change **only** pertains to the MSSP versions of these courses; and the regular versions will remain unchanged.}

CURRENT CROSS-LISTING DEPARTMENT/PROGRAM, if any:

TO BE OFFERED NEXT:  Sem./Year: _FALL__ / _2017__

INSTRUCTOR(S): MASANAO YAJIMA

DEPARTMENT CONTACT NAME AND POSITION: MARISA DISARNO, GRADUATE PROGRAM ADMINISTRATOR
DEPARTMENT CONTACT EMAIL AND PHONE: mdisarno@bu.edu, 617-353-2564

ITEMS PROPOSED FOR REVISION (check all that apply):

<table>
<thead>
<tr>
<th>Course Number</th>
<th>X 40 Word Description</th>
<th>X Prerequisites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-listing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

□ Divisional Studies Credit
□ Other (Explain)

Notes: The "short title" appears in the course inventory and on student transcripts and must be 15 characters
maximum including spaces. The “40 word description” appears in the CAS/GRS Bulletin.

PROPOSED REVISIONS: For each item checked above, provide the current information, then the proposed information, then a brief explanation for the proposed change, including the intended impact of the change.

1. [Course Number]
   a. Current information:  
      GRS MA 685
   b. Proposed information:  
      GRS MA 679
   c. Explanation & impact  
      New course number required
      (Please Note: 685 is run for both MSSP-only and non-MSSP audiences; this change **only** pertains to the MSSP versions of these courses; and the regular versions will remain unchanged.)

2. [Title]
   a. Current information:  
      Advanced Topics in Applied Statistical Analysis
   b. Proposed information:  
      Applied Statistical Machine Learning
   c. Explanation & impact:  
      More defined course names and new numbers will not only alleviate confusion with current students about course restrictions (MSSP only courses), but it will also be more effective for students entering industry to have more accurate and updated course names
      (Please Note: 685 is run for both MSSP-only and non-MSSP audiences; this change **only** pertains to the MSSP versions of these courses; and the regular versions will remain unchanged.)

3. [Short Title]
   a. Current information:  
      ADV STAT ANALY
   b. Proposed information:  
      APPL STAT LEARN
   c. Explanation & impact:  
      Change required to reflect new course title

4. [40 Word Description]
   a. Current information:  
      Continues topics of GRS MA 684 at a more advanced level. Canonical correlation, multivariate analysis of variance, multivariate regressions. Categorical dependent variables techniques; discriminant analysis, logistic regression, log-linear analysis. Factor analysis; principal-axes, rotations, factor scores. Cluster analysis. Power analysis. Extensive use of statistical software. This course
cannot be taken for credit in addition to the course with the same title that was previously numbered CAS MA 685.

b. Proposed information:
Continues topics of GRS MA 678 at a more advanced level. Application of supervised and unsupervised statistical machine learning techniques with extensive use of computation. Advanced topics such as analysis of network data, Bayesian nonparametric models will be considered.

c. Explanation & impact:
The original course was a topics course and our proposed list is an alternate set of topics. However, we are using the same textbook thus two thirds of the course material is entirely the same. The content of the nontrivial overlap that are advanced topics provided at the end of semester will be drawn mostly from the same text book that was recommended for the original class. *(Please Note: 685 is run for both MSSP-only and non-MSSP audiences; this change **only** pertains to the MSSP versions of these courses; and the regular versions will remain unchanged.)*

**IMPACT ON OTHER DEPARTMENTS/PROGRAMS:** Will any of these changes have an impact on students pursuing the degree requirements or expectations of other departments, programs, or schools?
Check one: □ Yes  X No

If YES, please identify impacts and attach cognate comment from the appropriate department/program/school.

**RESOURCE NEEDS: STAFFING, FACILITIES, AND EQUIPMENT:** As a result of the proposed changes, will there be any changes in the staffing, special facilities or equipment needs of the course (e.g. laboratory, library, instructional technology, technical resources, etc)?
Check one: □ Yes  X No

If YES, explain further and indicate whether currently available staffing, facilities, and equipment are adequate for the proposed course. *(NOTE: Approval of proposed revisions does not imply a change in resource commitments on the part of CAS.)*

**FURTHER INFORMATION THAT MUST BE SUBMITTED IN ORDER FOR THIS PROPOSAL TO BE CONSIDERED:**

1. A complete week-by-week SYLLABUS with student learning objectives, readings, and assignments that reflects the proposed changes (see guidelines on “Writing a Syllabus” on the Center for Teaching & Learning website. Be sure that syllabus includes your expectations for academic honesty, with URL for pertinent undergraduate or GRS academic conduct code(s).

2. Cognate comment from chairs or directors of relevant departments and/or programs. Use the form available here. You can consult with Joseph Bizup (CAS) or Jeffrey Hughes (GRS) to determine which departments or programs inside and outside of CAS would be appropriate.
DEPARTMENT APPROVAL:  

Other Department Chair(s) (for cross-listed courses)
DEAN'S OFFICE CURRICULUM ADMINISTRATOR USE ONLY

CAS/GRS CURRICULUM COMMITTEE APPROVAL:

☐ Approved    Date:____________________
☐ Tabled      Date:____________________
☐ Not Approved Date:____________________

Divisional Studies Credit:

☐ Endorsed
  ☐ HU
  ☐ MCS
  ☐ NS
  ☐ SS
☐ Not endorsed

________________________________________________________________________
Curriculum Committee Chair Signature and Date

Comments:

PROVISIONAL APPROVAL REQUESTED for Semester/Year __________________________

________________________________________________________________________
Dean of Arts & Sciences Signature and Date

Comments:

CAS FACULTY: Faculty Meeting Date: ____________________ ☐ Approved ☐ Not Approved

________________________________________________________________________
Curriculum Administrator Signature and Date

Comments:
Applied Statistical Machine Learning
GRS MA 679

Course Summary: Second in a two-semester sequence, continues topics of MA 686. Application of supervised and unsupervised statistical machine learning techniques with extensive use of computation. Advanced topics such as analysis of network data, Bayesian nonparametric models will be considered. These goals will be accomplished through an extensive use of statistical software.

Instructors: Masanao Yajima, PSY 229, yajima@bu.edu
Class Meetings: Tuesday/Thursday, 12:30 - 1:45pm, MCS B23
   Tuesday, 8:00 - 8:50pm, MCS B33 (Discussion Section)
Office Hour: Thursday, 3:00 - 4:30pm, and by appointment

Prereqs: Admission to the MS Program in Statistical Practice.

Course Description: The lecture meetings cover motivations, theoretical background and demonstrations of advanced methods in statistical learning. The lecture meetings are complemented by one discussion section per week, where exercises are discussed and the software implementation is reviewed in greater detail.

Instruction in Statistical Software and Computing: Knowledge of and comfort with statistical computing and software are, of course, critical to apply advanced statistical methods. Accumulating this knowledge is a life-long process and a skill in and of itself. To help you lay the foundation for that skill, in a formal manner, last semester the class MA 615 (Data Science in R) was available. This semester, you will build on that foundation with various R labs applying advanced methods of statistical learning.

Policies: Assignments will be posted to Blackboard with due dates/times and submission instructions. Overall, course policies are governed by Graduate School of Arts & Sciences (GRS). See GRS Policies, especially the GRS academic conduct code.

Grading: Grading for this course will be based on the following factors – reading quizzes, in class participation, assignments, a midterm exam and a final project. Each factor will be weighed 20%, 20%, 10%, 20% and 30%, respectively.

Resources: The class follows the textbook:
An Introduction to Statistical Learning with Applications in R.
by James, Witten, Hastie and Tibshirani (2013). Springer.

This textbook and corresponding R labs are available online: http://www-bcf.usc.edu/~gareth/ISL/. For a more advanced treatment of the topics covered by the lecture, the textbook can be combined with:
The Elements of Statistical Learning. Data Mining, Inference, and Prediction.
by Hastie, Tibshirani and Friedman (2009). Springer.

Which is also available online: https://statweb.stanford.edu/~tibs/ElemStatLearn/.
Note: This syllabus may change throughout the semester. Changes will be announced by email and the revision date will be updated.

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Discussion Section</th>
<th>Assignment Due</th>
<th>Reading / Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1 (Jan 19)</td>
<td>What is statistical learning?</td>
<td>N/A</td>
<td>N/A</td>
<td>ISL Ch 1, 2, 3</td>
</tr>
<tr>
<td>Week 2 (Jan 24)</td>
<td>GLM review</td>
<td>h2o glm</td>
<td>Reading quiz</td>
<td>ISL Ch 5, 6</td>
</tr>
<tr>
<td></td>
<td>Evaluating models</td>
<td></td>
<td></td>
<td>Regression homework</td>
</tr>
<tr>
<td>Week 3 (Jan 31)</td>
<td>Model selection &amp; regularization</td>
<td>regularization</td>
<td>Reading quiz</td>
<td>ISL Ch 7</td>
</tr>
<tr>
<td></td>
<td>Model selection &amp; regularization</td>
<td></td>
<td>Regression homework</td>
<td>Model selection &amp; regularization</td>
</tr>
<tr>
<td>Week 4 (Feb 7)</td>
<td>Regression &amp; Smoothing Splines</td>
<td>smoothing</td>
<td>Reading quiz</td>
<td>Smoothing homework</td>
</tr>
<tr>
<td></td>
<td>Regression &amp; Smoothing Splines</td>
<td></td>
<td></td>
<td>ISL Ch 4</td>
</tr>
<tr>
<td>Week 5 (Feb 14)</td>
<td>Classification</td>
<td>classification</td>
<td>Reading quiz</td>
<td>Classification homework</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td></td>
<td></td>
<td>ISL Ch 10.1 10.3</td>
</tr>
<tr>
<td>Week 6 (Feb 21)</td>
<td>No class</td>
<td></td>
<td></td>
<td>ISL Ch 8</td>
</tr>
<tr>
<td></td>
<td>Clustering</td>
<td></td>
<td>Classification homework</td>
<td>Clustering homework</td>
</tr>
<tr>
<td>Week 7 (Feb 28)</td>
<td>Tree-based Models</td>
<td>tree</td>
<td>Reading quiz</td>
<td>Tree-based Models homework</td>
</tr>
<tr>
<td></td>
<td>Tree-based Models</td>
<td></td>
<td></td>
<td>ISL Ch 9</td>
</tr>
<tr>
<td>Week 8 (Mar 7)</td>
<td>Spring break</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9 (Mar 14)</td>
<td>Support vector machines</td>
<td>svm</td>
<td>Reading quiz</td>
<td>SVM homework</td>
</tr>
<tr>
<td></td>
<td>Support vector machines</td>
<td></td>
<td></td>
<td>Tree-based Models homework ISL Ch 10.2 + extra</td>
</tr>
<tr>
<td>Week 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>(Mar 21)</td>
<td>Unsupervised learning</td>
<td>clustering, dimension reduction</td>
<td>Reading quiz: Unsupervised learning homework</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCA, Factor Analysis</td>
<td></td>
<td>Midterm Exam</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SVM homework</td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Working with network data (Eric)</td>
<td>network lab</td>
<td>Reading quiz</td>
<td></td>
</tr>
<tr>
<td>(Mar 28)</td>
<td></td>
<td></td>
<td>Unsupervised learning homework</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NN and Deep learning reading (TBD)</td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>Mixture Models and EM algorithm</td>
<td>EM and mixtures</td>
<td>Final Project</td>
<td></td>
</tr>
<tr>
<td>(Apr 4)</td>
<td></td>
<td></td>
<td>Reading (TBD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian Nonparametric Models</td>
<td>Bayesian nonparametrics</td>
<td></td>
</tr>
<tr>
<td>Week 13</td>
<td></td>
<td></td>
<td>Reading (TBD)</td>
<td></td>
</tr>
<tr>
<td>(Apr 11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 14</td>
<td>Neural Network and Deep Learning</td>
<td>deep learning</td>
<td>Reading (TBD)</td>
<td></td>
</tr>
<tr>
<td>(Apr 18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 15</td>
<td>TBD</td>
<td>TBD</td>
<td>Reading (TBD)</td>
<td></td>
</tr>
<tr>
<td>(Apr 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 16</td>
<td>Review</td>
<td>Final Project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(May 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>