

DEPARTMENT OF STATISTICS

Proposed New Course Offering

COURSE: STAT 760
Winter Quarter

The Elements of Statistical Learning

Statistical Learning methodology explores various ways of estimating functional dependencies between a response variable and possibly a large set of explanatory variables (features), when one is trying to find and understand an unknown, regular component within the realm of noisy, complex data. Modern regression and pattern recognition analyses as well as clustering and analysis of labeled and unlabelled data fall in this framework. Many regularized optimization problems can also be given a Bayesian interpretation. Briefly, statistical learning can be described as Machine Learning with a statistical (knowledge discovery) perspective.

Prerequisites: Probability/Math/Statistics - STAT 520/521 or higher, or EE 804/806, Familiarity with Matrix Algebra and Linear Regression Analysis, and Working Knowledge of a popular software package, such as Matlab, R, S-plus or SAS.

Instructor: TBA

Text Book:

Hastie, T., R. Tibshirani, and J. Friedman (2001) *The Elements of Statistical Learning - Data Mining, Inference and Prediction*, New York: Springer.

Supplementary Books and other Publications:

Recent Journal articles, lecture slides and other reference materials will be posted on the course website.

Grading: Grade will be based on an Individual/Team Project Report due at the end of the quarter, and the class presentation.

The weekly schedule is as follows:

Weeks 1-2: Search and Choose a data set/problem,

Week 3: Submit background description of the selected topic,

Week 4: Topic Finalized and approved,

Week 5: Project Description outline due,

Week 7: First draft of the Project Report due,

Week 9: Second draft of the Project Report due,

Week 10: Class Presentations

Exam Week: Class Presentations and Final Report Due

Tentative Lectures Schedule:

Week	Chapter	Topic
1	1-2	Overview of Supervised Learning
2	3	Linear Methods for Regression
3	4	Linear Methods for Classification
4	5	Basis Expansions and Regularization
5	6	Kernel Methods
6	7	Model Assessment and Selection
7	12	Support Vector Machines & Flexible Discriminants
8	9	Additive Models, Trees & Related Methods
9	10, 13	Boosting and Additive Trees, NN and Prototype Methods
10	14	Unsupervised Learning - Cluster Analysis`

B. General Information

1. Rational for the proposed course:

Recent emphasis on Data Mining and Machine Learning (DMML) in Computer Science, and the commonalities between DMML and statistical methods such as Classification, Regression, Clustering etc. has led to an interdisciplinary research area. This course provides the background material for Ph.D. students in Statistics. Of course, it will attract students from related areas.

2. Major/Minor Affected

Does not affect any current major/minor. But STAT and CSE faculty will be pursuing the creation of a graduate minor or a 'Graduate Interdisciplinary Specialization' in the area of Statistical and Machine Learning. This course will be required in that program.

3. Nature of program adjustment: There is full support from the Department of Statistics to teach this course to further enhance our Interdisciplinary Research profile. Pilot course(s) have been offered in the past three years. We have also hired a new faculty, starting Fall 05, with expertise and research interest in this area. Its regular offering does not require any budgetary adjustment.

8. Concurrence of other Academic Units

We have discussed this course offering on a regular basis with several faculty in Computer Science and Engineering (CSE) and Electrical Engineering (EE), who might be directly interested in its contents. In our pilot offerings of this material, as STAT 881 or STAT 894, students from both these departments were enrolled. In addition, several faculty and visitors audited the course. Profs. Srinivasan Parthasarathy and DeLiang Wang, from CSE have sent an expression of support (see attached e-mail messages).

From: "Parthasarathy Srinivasan" <srini@cse.ohio-state.edu>
To: <goel@stat.ohio-state.edu>
Subject: To Whom It May Concern
Date: Monday, June 20, 2005 12:23 PM

To Whom It May Concern:

I am in strong support of this course being proposed by Prem Goel and colleagues at Statistics. This course has been useful to CSE students in the past and will continue to be useful to CSE students in the data mining and machine learning areas.

Sincerely,

Srinivasan Parthasarathy, CSE, Assistant Professor

#

From: "DeLiang Wang" <dwang@cse.ohio-state.edu>
To: "Goel, Prem" <goel@stat.ohio-state.edu>
Cc: <dwang@cse.ohio-state.edu>
Subject: Re: Proposed course on Statistical Learning
Date: Monday, June 20, 2005 12:18 PM

TO WHOM IT MAY CONCERN,

Prof. Prem Goel of the Statistics Department shared with me the proposed course entitled "Elements of Statistical Learning". I support this course as it can be of benefit to students in Computer Science and Engineering (CSE) and it might become part of a GIS program currently under consideration.

There is substantial interest in machine learning and related statistical techniques within CSE, and we look forward to working with Prof. Goel and others to create a synergistic curriculum to serve the needs of interested OSU students.

Sincerely,

DeLiang Wang, Professor

--

Prof. DeLiang Wang
Department of Computer Science and Engineering
The Ohio State University
2015 Neil Ave.
Columbus, OH 43210-1277, U.S.A.

Phone: 614-292-6827 (OFFICE); 614-292-7402 (LAB)
Fax: 614-292-2911
URL: <http://www.cse.ohio-state.edu/~dwang>

#

From: "Parthasarathy Srinivasan" <srini@cse.ohio-state.edu>
To: <goel@stat.ohio-state.edu>
Subject: To Whom It May Concern
Date: Monday, June 20, 2005 12:23 PM

To Whom It May Concern:

I am in strong support of this course being proposed by Prem Goel and colleagues at Statistics. This course has been useful to CSE students in the past and will continue to be useful to CSE students in the data mining and machine learning areas.

Sincerely,

Srinivasan Parthasarathy, CSE, Assistant Professor

#

From: "DeLiang Wang" <dwang@cse.ohio-state.edu>
To: "Goel, Prem" <goel@stat.ohio-state.edu>
Cc: <dwang@cse.ohio-state.edu>
Subject: Re: Proposed course on Statistical Learning
Date: Monday, June 20, 2005 12:18 PM

TO WHOM IT MAY CONCERN,

Prof. Prem Goel of the Statistics Department shared with me the proposed course entitled "Elements of Statistical Learning". I support this course as it can be of benefit to students in Computer Science and Engineering (CSE) and it might become part of a GIS program currently under consideration.

There is substantial interest in machine learning and related statistical techniques within CSE, and we look forward to working with Prof. Goel and others to create a synergistic curriculum to serve the needs of interested OSU students.

Sincerely,

DeLiang Wang, Professor

--

Prof. DeLiang Wang
Department of Computer Science and Engineering
The Ohio State University
2015 Neil Ave.
Columbus, OH 43210-1277, U.S.A.

Phone: 614-292-6827 (OFFICE); 614-292-7402 (LAB)
Fax: 614-292-2911
URL: <http://www.cse.ohio-state.edu/~dwang>

#

DEPARTMENT OF STATISTICS

Proposed New Course Offering

COURSE: STAT 760

Winter Quarter

The Elements of Statistical Learning

Statistical Learning methodology explores various ways of estimating functional dependencies between a response variable and possibly a large set of explanatory variables (features), when one is trying to find and understand an unknown, regular component within the realm of noisy, complex data. Modern regression and pattern recognition analyses as well as clustering and analysis of labeled and unlabelled data fall in this framework. Many regularized optimization problems can also be given a Bayesian interpretation. Briefly, statistical learning can be described as Machine Learning with a statistical (knowledge discovery) perspective.

Prerequisites: Probability/Math/Statistics - STAT 520/521 or higher, or EE 804/806, Familiarity with Matrix Algebra and Linear Regression Analysis, and Working Knowledge of a popular software package, such as Matlab, R, S-plus or SAS.

Instructor: TBA

Text Book:

Hastie, T., R. Tibshirani, and J. Friedman (2001) *The Elements of Statistical Learning - Data Mining, Inference and Prediction*, New York: Springer.

Supplementary Books and other Publications:

Recent Journal articles, lecture slides and other reference materials will be posted on the course website.

Grading: Grade will be based on an Individual/Team Project Report due at the end of the quarter, and the class presentation.

The weekly schedule is as follows:

Weeks 1-2: Search and Choose a data set/problem,

Week 3: Submit background description of the selected topic,

Week 4: Topic Finalized and approved,

Week 5: Project Description outline due,

Week 7: First draft of the Project Report due,

Week 9: Second draft of the Project Report due,

Week 10: Class Presentations

Exam Week: Class Presentations and Final Report Due

Tentative Lectures Schedule:

Week	Chapter	Topic
1	1-2	Overview of Supervised Learning
2	3	Linear Methods for Regression
3	4	Linear Methods for Classification
4	5	Basis Expansions and Regularization
5	6	Kernel Methods
6	7	Model Assessment and Selection
7	12	Support Vector Machines & Flexible Discriminants
8	9	Additive Models, Trees & Related Methods
9	10, 13	Boosting and Additive Trees, NN and Prototype Methods
10	14	Unsupervised Learning - Cluster Analysis`

B. General Information

1. Rational for the proposed course:

Recent emphasis on Data Mining and Machine Learning (DMML) in Computer Science, and the commonalities between DMML and statistical methods such as Classification, Regression, Clustering etc. has led to an interdisciplinary research area. This course provides the background material for Ph.D. students in Statistics. Of course, it will attract students from related areas.

2. Major/Minor Affected

Does not affect any current major/minor. But STAT and CSE faculty will be pursuing the creation of a graduate minor or a 'Graduate Interdisciplinary Specialization' in the area of Statistical and Machine Learning. This course will be required in that program.

3. Nature of program adjustment: There is full support from the Department of Statistics to teach this course to further enhance our Interdisciplinary Research profile. Pilot course(s) have been offered in the past three years. We have also hired a new faculty, starting Fall 05, with expertise and research interest in this area. Its regular offering does not require any budgetary adjustment.

8. Concurrence of other Academic Units

We have discussed this course offering on a regular basis with several faculty in Computer Science and Engineering (CSE) and Electrical Engineering (EE), who might be directly interested in its contents. In our pilot offerings of this material, as STAT 881 or STAT 894, students from both these departments were enrolled. In addition, several faculty and visitors audited the course. Profs. Srinivasan Parthasarathy and DeLiang Wang, from CSE have sent an expression of support (see attached e-mail messages).

From: "DeLiang Wang" <dwang@cse.ohio-state.edu>
To: "Goel, Prem" <goel@stat.ohio-state.edu>
Cc: <dwang@cse.ohio-state.edu>
Subject: Re: Proposed course on Statistical Learning
Date: Monday, June 20, 2005 12:18 PM

TO WHOM IT MAY CONCERN,

Prof. Prem Goel of the Statistics Department shared with me the proposed course entitled "Elements of Statistical Learning". I support this course as it can be of benefit to students in Computer Science and Engineering (CSE) and it might become part of a GIS program currently under consideration.

There is substantial interest in machine learning and related statistical techniques within CSE, and we look forward to working with Prof. Goel and others to create a synergistic curriculum to serve the needs of interested OSU students.

Sincerely,

DeLiang Wang, Professor

--

Prof. DeLiang Wang
Department of Computer Science and Engineering
The Ohio State University
2015 Neil Ave.
Columbus, OH 43210-1277, U.S.A.

Phone: 614-292-6827 (OFFICE); 614-292-7402 (LAB)
Fax: 614-292-2911
URL: <http://www.cse.ohio-state.edu/~dwang>

#

From: "Parthasarathy Srinivasan" <srini@cse.ohio-state.edu>
To: <goel@stat.ohio-state.edu>
Subject: To Whom It May Concern
Date: Monday, June 20, 2005 12:23 PM

To Whom It May Concern:

I am in strong support of this course being proposed by Prem Goel and colleagues at Statistics. This course has been useful to CSE students in the past and will continue to be useful to CSE students in the data mining and machine learning areas.

Sincerely,

Srinivasan Parthasarathy, CSE, Assistant Professor

#