LECTURE 2 — EE660 — 8/23/2018

ANNOUNCEMENTS

1. HOMEWORK 1 POSTED - DUE TUE., 2:00PM.
2. MY OFFICE HOURS TOMORROW:
   12:00-1:00 PM (THIS WEEK ONLY)

TODAY'S LECTURE

1. KEY ISSUES AND CONCEPTS IN ML
2. COURSE OUTLINE.

KEY ISSUES AND CONCEPTS IN ML

1. HYPOTHESIS SET (MODELS BEING CONSIDERED)

E.G., 1D REGRESSION

MODELS: $\hat{f}_1(x) = w_0 + w_1 x$

$\hat{f}_2(x) = w_0 + w_1 x + w_2 x^2$.

OR MORE GENERALLY:

$$\hat{f}_d(x) = \sum_{i=0}^{d} w_i x^i, \quad 1 \leq d \leq d_{MAX}$$

OUR MODEL SELECTION AND LEARNING PROCESS

CHOOSES AMONG ALL d AND AMONG VALUES OF $\underline{w}$.

$\Rightarrow$ OUR HYPOTHESIS SET.

NOTE THAT A LIMITED HYPOTHESIS SET ENABLES GENERALIZATION TO NEW DATA.

$\rightarrow$ COMPLEXITY OF HYPOTHESIS SET IS IMPORTANT.

2. OBJECTIVE FUNCTION (FCN. BEING OPTIMIZED).

ALSO: "CRITERION" FCN.

Ex:

$$J(\underline{w}, \mathcal{D}) = MSE(\hat{y}, y_i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (\hat{y} - y_i)^2$$

IN WHICH: $\underbrace{\mathcal{D} = \{\underline{x}_i, y_i\}_{i=1}^{N}}$

J IS TO BE MINIMIZED.

3. OPTIMIZATION METHOD

MANY POSSIBILITIES

- NUMERIC AND ITERATIVE (CONTINUOUS)
  (grad. descent, Newton's method, ...)
- DISCRETE
- ALGEBRAIC
- STATISTICAL

>> CONVEX OR NON-CONVEX OBJECTIVE FCNS.

4. COMPLEXITY OF HYPOTHESIS SET, DATA, AND PROBLEM.

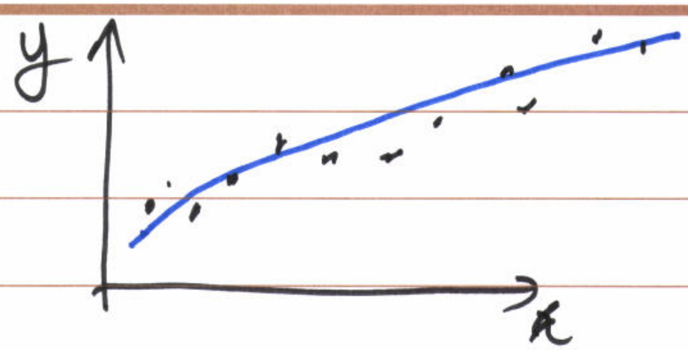
Ex: $\hat{f}_d(x) = \sum_{i=0}^{d} w_i x^i, \quad 1 \le d \le d_{MAX}.$

[Figs. 1.7, 1.18 in Murphy].

$\Rightarrow$ COMPLEXITY OF HYPOTHESIS SET, DATA, AND PROBLEM IS IMPORTANT.

5. ASSUMPTIONS AND PRIORS (INDUCTIVE BIAS)

e.g.:



A COMMON ASSUMPTION: "SMOOTHNESS" OF y BETWEEN NEARBY DATA POINTS IN FEATURE SPACE. (NOT ALWAYS VALID).

E.G.: RESTRICT COMPLEXITY OF HYPOTHESIS SET. (e.g., low d).

OR, USE AN APPROPRIATE PRIOR (⇒ REGULARIZER)

\5

EE 660 **Course Syllabus** Fall 2018

V1.0, 8/20/2018

Machine Learning from Signals: Foundations and Methods

Administrative information

Times and days

Lecture: TuTh 3:30 - 4:50 PM, OHE 122 and DEN@Viterbi

Discussion session: Wednesday 12:00 – 12:50 PM, OHE 122 and DEN@Viterbi

Catalogue description

Supervised, semi-supervised, and unsupervised machine learning; classification and regression. Model complexity, assessment, and selection; performance (error) on unseen data.

Course description

Foundations of machine learning, which apply to many or all algorithmic approaches, will be studied. These will include feasibility of learning; complexity of learning; regularization, overfitting and underfitting of models to data; model selection and assessment; and prediction of performance on unseen data. Particular methods that are key to machine learning from signals will also be covered. These will include linear and nonlinear techniques for regression as well as for classification, in the supervised learning realm. Also, methods described for classification by semi-supervised learning (using some labeled data and some unlabeled data for training), and for clustering by unsupervised learning (using only unlabeled data), will include statistical and distribution-free approaches. Feature selection, including the use of sparsity, will also be studied briefly. Students will be exposed to examples of techniques run on both synthetic and real-word data, through examples in lectures and the reading, as well as in homework problems and in the course project.

Learning Objectives

(1) To provide the student with a solid foundation in machine learning principles and the capability to apply them to problems.

(2) To give the student knowledge of common and successful methods (techniques and algorithms) in machine learning, and the ability to use them.

(3) To provide the student with sufficient foundation and knowledge so that he or she can learn about many of the plethora of machine learning techniques that now exist, on his or her own as needed.

6

Course Outline

There may be some minor changes in the topics or ordering. Number of lectures per topic is approximate.

intro

Introduction

1. Course introduction [Murphy] {1 lecture}

   Administrative information; introduction to the course and to machine learning

2. Key issues and concepts in machine learning. {1 lecture}

methods

Regression

3. Multidimensional regression [Murphy] {3 lectures}

   Linear regression, maximum-likelihood and MAP estimation, ridge regression, Bayesian regression. Learning linear and nonlinear relationships.

4. Logistic regression [Murphy] {1 lecture}

Foundations

Foundations of learning: Bayesian

5. Bayesian concept learning {1 lecture}

Foundations of learning: complexity

6. Feasibility of learning [AML] {1.5 lectures}

   Deterministic and statistical views; Hoeffding inequality (for bounding expected error on unlabeled data); inductive bias (model or data assumptions; e.g., parametric models, local smoothness)

7. Complexity of learning 1: generalization; estimation of error on new data; implications in dataset usage [AML] {3 lectures}

   Generalization bound, effective number of hypotheses, VC dimension, model complexity, sample complexity, dataset methodologies

8. Complexity of learning 2 [AML] {1.5 lectures}

   Bias-variance decomposition, learning curves, overfitting

Foundations & methods

Foundations and methods of learning: managing and controlling complexity

9. Regularization; feature reduction; sparsity [AML and Murphy] {3 lectures}

   Regularization as soft order constraints; Bayesian and MAP estimation for feature reduction; quadratic regularization; l_1 regularization, lasso, and sparsity; comparison of l_1 and l_2 regularizers; nonconvex regularizers, l_0 regularization, and bridge regression

10. Model selection [AML and Murphy] {1 lecture}

    Model selection and validation

methods

Graphical and nonlinear methods of learning

11. Boosting techniques and decision trees [Murphy] {3 lectures}

    Adaptive basis models; classification and regression trees (CART); random forests; boosting (Adaboost).

12. Kernel methods (theory and practice) [Murphy] {1 lecture}

    Examples of kernels (radial basis function, Mercer), kernel machines; kernel trick. Examples of support vector machine variants.

Semi-supervised and unsupervised learning methods

13. Semi-supervised learning for classification [Zhu] {2 lectures}

    Overview, including inductive vs. transductive semi-supervised learning; mixture models and Expectation Maximization for semi-supervised learning.

14. Unsupervised learning for clustering: statistical techniques [Xu] {1 lecture}

    Statistical techniques including mixture models; Maximum Likelihood; Expectation Maximization

15. Unsupervised learning for clustering: other techniques [Murphy and Xu] {2 lectures}

    Similarity measures; evaluating clustering quality and choosing K; hierarchical and graph clustering (agglomerative, divisive, Bayesian)*

Other topics*

16. Selected topic(s) of student interest. {1.5 lectures}

    A list of topics will be generated by suggestion and discussion. Topics to be covered will be chosen by discussion and vote.

* *As time permits.*