Machine Learning Algorithms 1
module 4, academic year 2021–2022

Sergei Golovan
New Economic School
sgolovan@nes.ru

TA: TBD

Course description

The course "Machine learning algorithms 1" is designed to further develop students' skill in statistical, econometric, and programming tools which are widely used in economics, particularly in finance and macroeconomics. The course topic is the statistical learning from data, the field which is especially useful for applications in the fields where large amount of data can be collected, e.g. in finance. The course is an elective one. It consists of 14 lectures and 7 seminars.

Course requirements, grading, and attendance policies

The course uses statistical technique developed through the required courses in Econometrics. Other than that, it doesn't have any special prerequisites except for the standard calculus, linear algebra, and probability courses.

There will be 4 home assignments which will constitute 40% of the final grade. The final exam will account for the remaining 60%.

Course contents

1. Introduction to statistical learning
   (a) What is statistical learning?
   (b) Supervised and unsupervised learning
   (c) Regression and classification
2. Supervised learning
   (a) Linear regression (regularization: ridge regression, lasso)
   (b) Linear classification (discriminant analysis, logistic models)
   (c) Polynomial and non-parametric models
   (d) Additive models
   (e) Tree-based methods
   (f) Neural networks

   (g) Support vector machines
   (h) Nearest neighbor classification

3. Unsupervised learning
   (a) Association rules
   (b) Cluster analysis
   (c) Factor analysis

Description of course methodology

Lectures will proceed from motivating examples and sample models in economics to general principles of statistical modeling. Also, a number of computer exercises will be distributed in order to give students an opportunity to practice the statistical techniques.

Sample tasks for course evaluation

1. Suppose that we take a data set, divide it into equally-sized training and test sets, and then try out two different classification procedures. First we use logistic regression and get an error rate of 20% on the training data and 30% on the test data. Next we use 1-nearest neighbors (i.e. $K = 1$) and get an average error rate (averaged over both test and training data sets) of 18%. Based on these results, which method should we prefer to use for classification of new observations? Why?

2. Suppose we produce ten bootstrapped samples from a data set containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of X, produce 10 estimates of $\mathrm{P}(\text{Class is Red} \mid X)$:

$$0.1,\ 0.15,\ 0.2,\ 0.2,\ 0.55,\ 0.6,\ 0.6,\ 0.65,\ 0.7,\ \text{and } 0.75.$$

   There are two common ways to combine these results together into a single class prediction. One is the majority vote approach. The second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?

3. Here we explore the maximal margin classifier on a toy data set.
   (a) We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label.

Obs.	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

   Sketch the observations.

(b) Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.

(c) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$, and classify to Blue otherwise." Provide the values for β_0, β_1, and β_2.

(d) On your sketch, indicate the margin for the maximal margin hyperplane.

(e) Indicate the support vectors for the maximal margin classifier.

(f) Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.

(g) Sketch a hyperplane that is not the optimal separating hyperplane, and provide the equation for this hyperplane.

(h) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

4. Consider a neural network for a K class outcome that uses crossentropy loss. If the network has no hidden layer, show that the model is equivalent to the multinomial logistic model.

5. In this problem, you will perform K-means clustering manually, with $K = 2$, on a small example with $n = 6$ observations and $p = 2$ features. The observations are as follows.

Obs.	X_1	X_2
1	1	4
2	1	3
3	0	4
4	5	1
5	6	2
6	4	0

(a) Plot the observations.

(b) Randomly assign a cluster label to each observation. Report the cluster labels for each observation.

(c) Compute the centroid for each cluster.

(d) Assign each observation to the centroid to which it is closest, in terms of Euclidean distance. Report the cluster labels for each observation.

(e) Repeat (5c) and (5d) until the answers obtained stop changing.

(f) In your plot from (5a), color the observations according to the cluster labels obtained.

Course materials

1. James G., Witten D., Hastie T., Tibshiriani R. (2015) An introduction to statistical learning with applications in R, 6th edition, Springer.

2. Hastie T., Tibshiriani R., Friedman J. (2008) The elements of statistical learning. Datamining, inference, and prediction, 2nd edition, Springer.

3. Witten I. H., Frank E. (2005) Data mining. Practical machine learning tools and techniques, 2nd edition, Morgan Kaufman.

Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.