Tue Herlau, Mikkel N. Schmidt and Morten Mørup

Introduction to Machine Learning and Data Mining

Material for continuing education course, Spring 2019

February 18, 2019

Technical University of Denmark

Notation cheat sheet

	Matlab var.	Type	Size	Description
	`X`	Numeric	$N \times M$	Data matrix: The rows correspond to N data objects, each of which contains M attributes.
	`attributeNames`	Cell array	$M \times 1$	Attribute names: Name (string) for each of the M attributes.
	`N`	Numeric	Scalar	Number of data objects.
	`M`	Numeric	Scalar	Number of attributes.
Regression	`y`	Numeric	$N \times 1$	Dependent variable (output): For each data object, `y` contains an output value that we wish to predict.
Classification	`y`	Numeric	$N \times 1$	Class index: For each data object, `y` contains a class index, $\mathtt{y}_n \in \{0, 1, \ldots, C-1\}$, where C is the total number of classes.
	`classNames`	Cell array	$C \times 1$	Class names: Name (string) for each of the C classes.
	`C`	Numeric	Scalar	Number of classes.
Cross-validation				All variables mentioned above appended with `_train` or `_test` represent the corresponding variable for the training or test set.
	`⋆_train`	—	—	Training data.
	`⋆_test`	—	—	Test data.

This book attempts to give a concise introduction to machine-learning concepts. We believe this is best accomplished by clearly stating what a given method actually does as a sequence of mathematical operations, and use illustrations and text to provide an intuition. We will therefore make use of tools from linear algebra, probability theory and analysis to describe the methods, focusing on using as small a set of concepts as possible and strive towards maximal consistency.

In the following, vectors will be denoted by lower-case roman letters $\boldsymbol{x}, \boldsymbol{y}, \ldots$ and matrices by bolder, upper case roman letters $\boldsymbol{A}, \boldsymbol{B}, \ldots$. A superscript T denote the transpose. For instance

$$\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 1 & -2 \end{bmatrix} \text{ and if } \boldsymbol{x} = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix} \text{ then } \boldsymbol{x}^T = \begin{bmatrix} -1 & 4 & 1 \end{bmatrix}.$$

The ith element of a vector is written as x_i and the i,j'th element of a matrix as A_{ij} (and sometimes $A_{i,j}$ to avoid ambiguity). In the preceding example, $x_2 = 4$ and $A_{2,3} = -2$. During this course the observed data set, which we feed into our machine learning methods, will consist of N observations where each observation consist of a M dimensional vector. For instance if we have N observations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ then any given observation will consist of M numbers:

$$\boldsymbol{x} = \begin{bmatrix} x_1 & \ldots, x_M \end{bmatrix}^T.$$

For convenience, we will often combine the observations into an $N \times M$ data matrix $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T \end{bmatrix}$$

in which the ith row of $\boldsymbol{X}$ corresponds to the row vector $\boldsymbol{x}_i^T$. We will use this notation for our data matrix and the *rows* of $\boldsymbol{X}$ will correspond to N *observations* and the M *columns* of $\boldsymbol{X}$ will correspond to M *attributes*. Often each of the observations $\boldsymbol{x}_i$ will come with a *label* or *target* y_i corresponding to a feature of $\boldsymbol{x}_i$ which we are interested in predicting. In this case we will collect the labels in a N-dimensional vector $\boldsymbol{y}$ and the pair $(\boldsymbol{X}, \boldsymbol{y})$ will be all the data available for the machine learning method. A more comprehensive translation of the notation as used in this book and in the exercises can be found in the table on the previous page. Finally, the reader should be familiar with the big-sigma notation which allows us to conveniently write sums and products of multiple terms:

$$\sum_{i=1}^{n} f(i) = f(1) + f(2) + \cdots + f(n-1) + f(n)$$
$$\prod_{i=1}^{n} f(i) = f(1) \times f(2) \times \cdots \times f(n-1) \times f(n).$$

As an example, if $f(i) = i^2$ and $n = 4$ we have

$$\sum_{i=1}^{4} f(i) = 1^2 + 2^2 + 3^2 + 4^2 = 30, \quad \prod_{i=1}^{4} f(i) = 1^2 \times 2^2 \times 3^2 \times 4^2 = 576.$$

Contents

Part II Supervised learning

Part III Unsupervised learning