REPORT: PATTERN RECOGNITION

Leif Van Holland

University of Bonn

1. INTRODUCTION

In the last decades, unprecedented amounts of data are collected in various places throughout the analog and digital world. The amount of information renders manual interpretation unfeasible if not impossible. Therefore the topic of *pattern recognition* plays an increasingly important role in modern technology.
The university lecture "Pattern Recognition"...
In the following... in order of the projects

2. REGRESSION

The term *regression* refers to a statistical method of estimating relationships between variables. Given a set of data $D = \{(x_i, y_i)\}_{i=1}^N$, where $x_i \in \mathbb{R}^N$ and $y_i \in \mathbb{R}$, the goal is to find a set of parameters $\theta \in \mathbb{R}^k$ of a given model $y : \mathbb{R}^N \times \mathbb{R}^k \to \mathbb{R}$, such that y *predicts* the values y_i based on x_i as an input with minimal error. In other words, we look for parameter values $\hat{\theta}$ such that

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, E(\theta). \tag{1}$$

E denotes the *objective function* that depends on the problem at hand. Most times however, E measures a distance between the target output y_i and the model prediction $y(x_i, \hat{\theta})$.

2.1. Maximum Likelihood Estimation

If we choose a probability distribution N based on parameters θ as our model, i.e. we suspect our data to be realizations of random variables $X_i \sim N[\theta]$, we can calculate the probability of any possible realization x_i of X_i depending on θ. Bayes' formula yields

$$P(x_i|\theta) = \frac{P(\theta|x_i)P(x_i)}{P(\theta)}$$

and we define the *likelihood* L that parameters θ are responsible for generating the set of data D as a a function

$$L(\theta, D) := P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}.$$

If we assume that the random variables X_i are i.i.d., we can further simplify that definition and rewrite the joint density $P(D)$ as a product and thus get

$$L(\theta, D) = \prod_{x_i \in D} P(\theta|x_i).$$

For numerical stability, one often considers the *log-likelihood* function

$$\mathcal{L}(\theta, D) = \ln L(\theta, D) = \sum_{x_i \in D} \ln P(\theta|x_i).$$

To find the parameters θ that are most likely to have generated D, we look for the *maximum likelihood estimate*

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta, D) = \underset{\theta}{\operatorname{argmin}} -L(\theta, D)$$

2.2. Normal and Weibull Distribution

In the first project we looked at two distributions applied to 1-dimensional data. The **1D normal distribution** is $N[\mu, \sigma^2]$ depending on two parameters $\mu, \sigma^2 \in \mathbb{R}$ with the density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

Using the method of maximum likelihood estimation we can specify the optimal choice $\hat{\mu}$ and $\hat{\sigma}^2$ for both parameters of the normal distribution directly, which coincide with the *sample mean* and *population variance* of the given set of data respectively:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

The result of this estimation on the body sizes from the data introduced in **??** can be seen in figure 1.

The **Weibull distribution** on the other hand uses parameters $\kappa, \alpha \in \mathbb{R}$ and is defined by the density

$$f(x) = \frac{\kappa}{\alpha}\left(\frac{x}{\alpha}\right)^{\kappa-1} e^{-(\frac{x}{\alpha})^\kappa}.$$

Estimating the parameters here is more involved, as there is no closed-form solution for the estimates $\hat{\kappa}$ and $\hat{\alpha}$. A maximum of the log-likelihood function

$$L(\alpha, \kappa) = N(\log\kappa - \kappa\log\alpha) + (\kappa - 1)\sum_i \log d_i - \sum_i \left(\frac{d_i}{\alpha}\right)^\kappa$$

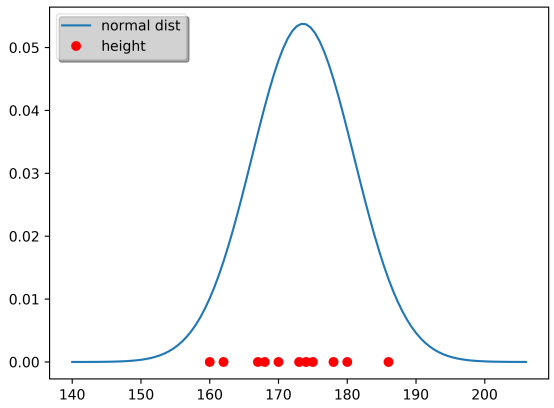


Fig. 1. Resulting PDF of a normal distribution maximizing the likelihood w.r.t. the body sizes from the data set introduced in **??**.

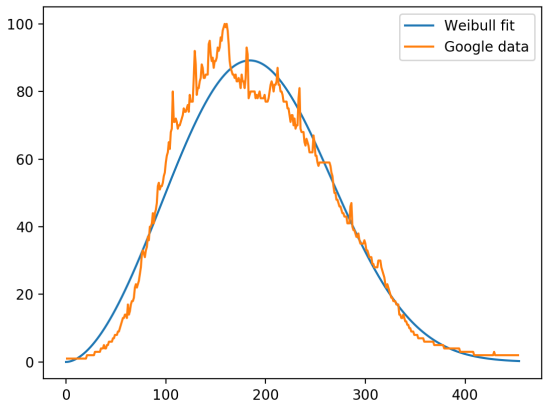


Fig. 2. MLE-fitted PDF of a Weibull distribution on Google Trends data describing the evolution of global interest in the search term *"myspace"*.

has therefore to be found numerically and we used Newton's method initialized with $\kappa = 1$ and $\alpha = 1$. To optimize the calculation time we rewrote the log-likelihood using a histogram $h(x_j) = h_j$ for all occurring values x_j in the data, which reduces the number of elements the sums in L are iterating over. $L(\alpha, \kappa)$ then equals the term

$$N(\log\kappa - \kappa\log\alpha) + (\kappa - 1)\sum_j x_j \log h_j - \sum_j \left(\frac{x_j \cdot h_j}{\alpha}\right)^\kappa$$

We used the described method on Google Trends data about global interest in the search term *"myspace"*, measured every week between January 1, 2003 and March 16, 2012. Figure 2 shows that assuming a Weibull distribution reasonably describes the overall trend for the most part, although one can clearly observe a deviation from the actual graph between weeks 100 and 200.

2.3. Least Squares Technique

A different but also very common technique is to formulate the problem as a distance $\|Xw - y\|_2$ between a design matrix $X \in \mathbb{N} \times$ times the solution vector $w \in \mathbb{R}^d$ and a target vector $y \in \mathbb{N}$ where

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

This distance is to be minimized, so as an objective function we choose

$$E(w) = \|Xw - y\|_2^2$$

whereby the norm is often squared for simplicity. basic idea, normal equations, pseudoinverse, least squares as special case of MLE (under gaussian noise) example: fractal dimensions, polynomials

2.4. Bayesian Regression

first conditional expectation, this leads to bayesian regression

2.5. Boolean Functions

example to use least squares, show how we can find a linear combination of the boolean basis to reconstruct the rules. In fact we increase the dimension of the problem to get better results.

3. CLASSIFICATION

basic idea of classification

3.1. Nearest Neighbors

short explanation of nn

3.2. kD-Trees

short explanation of kd trees

3.3. k-Means Clustering

different algorithms for k-means

3.4. Spectral Clustering

stuff about fiedler vector, spectral clustering from computer vision

3.5. Dimensionality Reduction

pca and lda, explanation of differences

3.6. Non-Monotonous Neurons

3.7. support vector machines

short explanation of kernel trick and its implementation

4. NUMERICAL INSTABILITIES

throughout the projects special care for numerical issues

5. CONCLUSION

6. MODEL FITTING

Most of the problems considered in this report are based on a given set of data $D = \{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. The general goal is to find a set of parameters $\theta \in \mathbb{R}^k$ of a given model $y : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, such that y *predicts* the values y_i based on x_i as an input with minimal error. In other words, we look for parameter values $\hat{\theta}$ so that

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, E(\theta) \tag{2}$$

E denotes the *objective function* that depends on the problem at hand. Most times however, E measures a distance between the target output y_i and the model prediction $y(x_i, \hat{\theta})$.

6.1. Least squares

A common approach is to construct an overdetermined linear system of m equations and n variables $Ax = b$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are given and we are looking for the solution $x \in \mathbb{R}^n$. Since it is overdetermined, which implies $m \geq n$, there is no exact solution in general. We rather look for an x that minimizes the (euclidean) length of the residual $r = b - Ax \in \mathbb{R}^m$, i.e.

$$\|b - Ax\|_2 \to \min \tag{3}$$

Minimizing the length of r is equivalent to minimizing the squares of the entries of the vector r. Therefore this formulation is called the *least squares* problem.
There is a general solution for this problem: One can show that if and only if the matrix A has full rank, the *normal equations*

$$A^* Ax = A^* b \tag{4}$$

hold for any solution residual-minimizing x. This solution is in fact unique, as least squares is a convex problem. Equation (4) motivates the definition of the *pseudoinverse* of A

$$A^+ = (A^T A)^{-1} A^T \in \mathbb{R}^{n \times m} \tag{5}$$

that provides the unique solution $x = A^+ b$ [1, 2].

6.2. Regression using least squares

7. PAGE TITLE SECTION

The paper title (on the first page) should begin 1.38 inches (35 mm) from the top edge of the page, centered, completely capitalized, and in Times 14-point, boldface type. The authors' name(s) and affiliation(s) appear below the title in capital and lower case letters. Papers with multiple authors and affiliations may require two or more lines for this information. Please note that papers should not be submitted blind; include the authors' names on the PDF.

8. TYPE-STYLE AND FONTS

To achieve the best rendering both in printed proceedings and electronic proceedings, we strongly encourage you to use Times-Roman font. In addition, this will give the proceedings a more uniform look. Use a font that is no smaller than nine point type throughout the paper, including figure captions.

In nine point type font, capital letters are 2 mm high. **If you use the smallest point size, there should be no more than 3.2 lines/cm (8 lines/inch) vertically.** This is a minimum spacing; 2.75 lines/cm (7 lines/inch) will make the paper much more readable. Larger type sizes require correspondingly larger vertical spacing. Please do not double-space your paper. TrueType or Postscript Type 1 fonts are preferred.

The first paragraph in each section should not be indented, but all the following paragraphs within the section should be indented as these paragraphs demonstrate.

9. MAJOR HEADINGS

Major headings, for example, "1. Introduction", should appear in all capital letters, bold face if possible, centered in the column, with one blank line before, and one blank line after. Use a period (".") after the heading number, not a colon.

9.1. Subheadings

Subheadings should appear in lower case (initial word capitalized) in boldface. They should start at the left margin on a separate line.

9.1.1. Sub-subheadings

Sub-subheadings, as in this paragraph, are discouraged. However, if you must use them, they should appear in lower case (initial word capitalized) and start at the left margin on a separate line, with paragraph text beginning on the following line. They should be in italics.

10. PRINTING YOUR PAPER

Print your properly formatted text on high-quality, 8.5 x 11-inch white printer paper. A4 paper is also acceptable, but please leave the extra 0.5 inch (12 mm) empty at the BOTTOM of the page and follow the top and left margins as specified. If the last page of your paper is only partially filled, arrange the columns so that they are evenly balanced if possible, rather than having one long column.

In LaTeX, to start a new column (but not a new page) and help balance the last-page column lengths, you can use the command "\pagebreak" as demonstrated on this page (see the LaTeX source below).

11. PAGE NUMBERING

Please do **not** paginate your paper. Page numbers, session numbers, and conference identification will be inserted when the paper is included in the proceedings.

12. ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. Colors may be used, but they should be selected so as to be readable when printed on a black-only printer.

Since there are many ways, often incompatible, of including images (e.g., with experimental results) in a LaTeX document, below is an example of how to do this [3].

13. FOOTNOTES

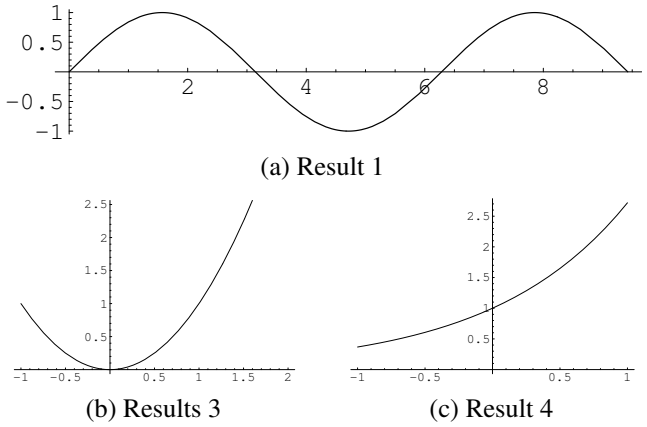
Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

14. COPYRIGHT FORMS

You must include your fully completed, signed IEEE copyright release form when form when you submit your paper. We **must** have this form before your paper can be published in the proceedings.

15. REFERENCES

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [4]. An additional final page (the fifth page, in most cases) is allowed, but must contain only references to the prior literature.



(a) Result 1

(b) Results 3

(c) Result 4

Fig. 3. Example of placing a figure with experimental results.

16. REFERENCES

[1] Lloyd N Trefethen and David Bau III, *Numerical linear algebra*, vol. 50, Siam, 1997.

[2] Prof. Dr. Christian Bauckhage, "lectures on pattern recognition," 2017, notes can be found at http://www.researchgate.net/project/lectures-on-pattern-recognition, accessed on 2018-02-20.

[3] A.B. Smith, C.D. Jones, and E.F. Roberts, "Article title," *Journal*, vol. 62, pp. 291–294, January 1920.

[4] C.D. Jones, A.B. Smith, and E.F. Roberts, "Article title," in *Proceedings Title*. IEEE, 2003, vol. II, pp. 803–806.