

# Analysis of Logistic Regression and Linear Discriminant Analysis for Breast Cancer Diagnosis and Red Wine Quality

Tianzhu Fu, Xingyu Li, and Wenzong Xia

(Dated: 28 September 2019)

**Abstract:** This work applies logistic regression using gradient descent and linear discriminant analysis to the tasks of predicting breast cancer tumor types and red wine quality. The analytic methods of logistic regression using gradient descent and linear discriminant analysis are compared and it is shown that linear discriminant analysis is more effective and efficient for binary classification tasks. We also demonstrate how gradient descent hyperparameters affect convergence rate and the quality of the obtained results.

## I. INTRODUCTION

### Preliminaries

Logistic regression is a common regression analysis to conduct when the dependent variable is dichotomous (binary).

### A. Second-level heading: Formatting

This file may be formatted in both the `preprint` (the default) and `reprint` styles; the latter format may be used to mimic final journal output. Either format may be used for submission purposes; however, for peer review and production, AIP will format the article using the `preprint` class option. Hence, it is essential that authors check that their manuscripts format acceptably under `preprint`. Manuscripts submitted to AIP that do not format correctly under the `preprint` option may be delayed in both the editorial and production processes.

The `widetext` environment will make the text the width of the full page, as on page 2. (Note the use the `\pageref{#1}` to get the page number right automatically.) The width-changing commands only take effect in `twocolumn` formatting. It has no effect if `preprint` formatting is chosen instead.

### 1. Third-level heading: Citations and Footnotes

Citations in text refer to entries in the Bibliography; they use the commands `\cite{#1}` or `\onlinecite{#1}`. Because REVTeX uses the `natbib` package of Patrick Daly, its entire repertoire of commands are available in your document; see the `natbib` documentation for further details. The argument of `\cite` is a comma-separated list of *keys*; a key may consist of letters and numerals.

By default, citations are numerical;<sup>?</sup> author-year citations are an option. To give a textual citation, use `\onlinecite{#1}`: (Refs. ? ? ? ). REVTeX “collapses” lists of consecutive numerical citations when appropriate. REVTeX provides the ability to properly punctuate textual citations in author-year style; this facility works cor-

rectly with numerical citations only with `natbib`’s `compress` option turned off. To illustrate, we cite several together<sup>? ? ? ?</sup>, and once again (Refs. ? ? ? ? ). Note that, when numerical citations are used, the references were sorted into the same order they appear in the bibliography.

A reference within the bibliography is specified with a `\bibitem{#1}` command, where the argument is the citation key mentioned above. `\bibitem{#1}` commands may be crafted by hand or, preferably, generated by using BibTeX. The AIP styles for REVTeX 4 include BibTeX style files `aipnum.bst` and `aipauth.bst`, appropriate for numbered and author-year bibliographies, respectively. REVTeX 4 will automatically choose the style appropriate for the document’s selected class options: the default is numerical, and you obtain the author-year style by specifying a class option of `author-year`.

This sample file demonstrates a simple use of BibTeX via a `\bibliography` command referencing the `aipsamp.bib` file. Running BibTeX (in this case `bibtex aipsamp`) after the first pass of L<sup>A</sup>T<sub>E</sub>X produces the file `aipsamp.bbl` which contains the automatically formatted `\bibitem` commands (including extra markup information via `\bibinfo` commands). If not using BibTeX, the `thebibliography` environment should be used instead.

*a. Fourth-level heading is run in.* Footnotes are produced using the `\footnote{#1}` command. Numerical style citations put footnotes into the bibliography<sup>1</sup>. Author-year and numerical author-year citation styles (each for its own reason) cannot use this method. Note: due to the method used to place footnotes in the bibliography, *you must re-run BibTeX every time you change any of your document’s footnotes.*

## II. MATH AND EQUATIONS

Inline math may be typeset using the `$` delimiters. Bold math symbols may be achieved using the `bm` package and the `\bm{#1}` command it supplies. For instance, a bold  $\alpha$  can be typeset as `$\bm{\alpha}$` giving  $\alpha$ . Fraktur and Blackboard (or open face or double struck) characters should be typeset using the `\mathfrak{#1}` and `\mathbb{#1}` commands respectively. Both are supplied by the `amssymb` package. For example, `$\mathbb{R}$`

gives  $\mathbb{R}$  and  $\mathfrak{G}$  gives  $\mathfrak{G}$

In L<sup>A</sup>T<sub>E</sub>X there are many different ways to display equations, and a few preferred ways are noted below. Displayed math will center by default. Use the class option `fleqn` to flush equations left.

Below we have numbered single-line equations, the most common kind:

$$\chi_+(p) \lesssim [2|\mathbf{p}|(|\mathbf{p}| + p_z)]^{-1/2} \begin{pmatrix} |\mathbf{p}| + p_z \\ px + ip_y \end{pmatrix}, \quad (1)$$

$$\left\{ 1234567890abc123\alpha\beta\gamma\delta1234556\alpha\beta \frac{1\sum_b^a}{A^2} \right\}. \quad (2)$$

Note the open one in Eq. (2).

Not all numbered equations will fit within a narrow column this way. The equation number will move down automatically if it cannot fit on the same line with a one-line equation:

$$\left\{ ab12345678abc123456abcdef\alpha\beta\gamma\delta1234556\alpha\beta \frac{1\sum_b^a}{A^2} \right\}. \quad (3)$$

When the `\label{#1}` command is used [cf. input for Eq. (2)], the equation can be referred to in text without knowing the equation number that T<sub>E</sub>X will assign to it. Just use `\ref{#1}`, where `#1` is the same name that used in the `\label{#1}` command.

Unnumbered single-line equations can be typeset using the `[, \]` format:

$$g^+g^+ \rightarrow g^+g^+g^+g^+ \dots, \quad q^+q^+ \rightarrow q^+g^+g^+ \dots$$

### A. Multiline equations

Multiline equations are obtained by using the `eqnarray` environment. Use the `\nonumber` command at the end of each line to avoid assigning a number:

$$\mathcal{M} = ig_Z^2 (4E_1 E_2)^{1/2} (l_i^2)^{-1} \delta_{\sigma_1, -\sigma_2} (g_{\sigma_2}^e)^2 \chi_{-\sigma_2}(p_2) \times [\epsilon_j l_i \epsilon_i]_{\sigma_1} \chi_{\sigma_1}(p_1), \quad (4)$$

$$\sum |M_g^{\text{viol}}|^2 = g_S^{2n-4} (Q^2)^{N^{n-2}} (N^2 - 1) \times \left( \sum_{i < j} \right) \sum_{\text{perm}} \frac{1}{S_{12}} \frac{1}{S_{12}} \sum_{\tau} c_{\tau}^f. \quad (5)$$

**Note:** Do not use `\label{#1}` on a line of a multiline equation if `\nonumber` is also used on that line. Incorrect cross-referencing will result. Notice the use `\text{#1}` for using a Roman font within a math environment.

To set a multiline equation without *any* equation numbers, use the `\begin{eqnarray*}`, `\end{eqnarray*}` format:

$$\sum |M_g^{\text{viol}}|^2 = g_S^{2n-4} (Q^2)^{N^{n-2}} (N^2 - 1) \times \left( \sum_{i < j} \right) \left( \sum_{\text{perm}} \frac{1}{S_{12} S_{23} S_{n1}} \right) \frac{1}{S_{12}}.$$

To obtain numbers not normally produced by the automatic numbering, use the `\tag{#1}` command, where `#1` is the desired equation number. For example, to get an equation number of (2.6'),

$$g^+g^+ \rightarrow g^+g^+g^+g^+ \dots, \quad q^+q^+ \rightarrow q^+g^+g^+ \dots \quad (2.6')$$

A few notes on `\tag{#1}`. `\tag{#1}` requires `amsmath`. The `\tag{#1}` must come before the `\label{#1}`, if any. The numbering set with `\tag{#1}` is *transparent* to the automatic numbering in REV<sub>T</sub>E<sub>X</sub>; therefore, the number must be known ahead of time, and it must be manually adjusted if other equations are added. `\tag{#1}` works with both single-line and multiline equations. `\tag{#1}` should only be used in exceptional case - do not use it to number all equations in a paper.

Enclosing single-line and multiline equations in `\begin{subequations}` and `\end{subequations}` will produce a set of equations that are “numbered” with letters, as shown in Eqs. (6a) and (6b) below:

$$\left\{ abc123456abcdef\alpha\beta\gamma\delta1234556\alpha\beta \frac{1\sum_b^a}{A^2} \right\}, \quad (6a)$$

$$\mathcal{M} = ig_Z^2 (4E_1 E_2)^{1/2} (l_i^2)^{-1} (g_{\sigma_2}^e)^2 \chi_{-\sigma_2}(p_2) \times [\epsilon_i]_{\sigma_1} \chi_{\sigma_1}(p_1). \quad (6b)$$

Putting a `\label{#1}` command right after the `\begin{subequations}`, allows one to reference all the equations in a subequations environment. For example, the equations in the preceding subequations environment were Eqs. (6).

### 1. Wide equations

The equation that follows is set in a wide format, i.e., it spans across the full page. The wide format is reserved for long equations that cannot be easily broken into four lines or less:

$$\mathcal{R}^{(d)} = g_{\sigma_2}^e \left( \frac{[\Gamma^Z(3, 21)]_{\sigma_1}}{Q_{12}^2 - M_W^2} + \frac{[\Gamma^Z(13, 2)]_{\sigma_1}}{Q_{13}^2 - M_W^2} \right) + x_W Q_e \left( \frac{[\Gamma^\gamma(3, 21)]_{\sigma_1}}{Q_{12}^2 - M_W^2} + \frac{[\Gamma^\gamma(13, 2)]_{\sigma_1}}{Q_{13}^2 - M_W^2} \right). \quad (7)$$

This is typed to show the output is in wide format. (Since there is no input line between `\equation` and this paragraph, there is no paragraph indent for this paragraph.)

<sup>1</sup>Automatically placing footnotes into the bibliography requires using BibTeX to compile the bibliography.