Local Outlier Factor

Der Local Outlier Factor (LOF, etwa "Lokaler Ausreißerfaktor") ist ein Algorithmus zur Erkennung von dichtebasierten Ausreißern von Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng und Jörg Sander. Die Kernidee von LOF besteht darin, die Dichte eines Punktes mit den Dichten seiner Nachbarn zu vergleichen. Ein Punkt, der "dichter" ist als seine Nachbarn, befindet sich in einem Cluster. Ein Punkt, mit einer deutlich geringeren Dichte als seine Nachbarn, ist hingegen ein Ausreißer.

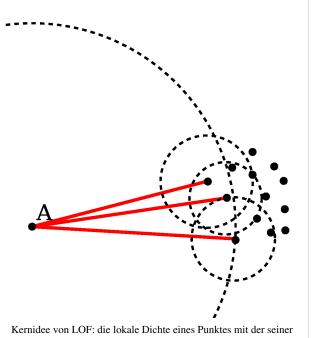
LOF hat viele Konzepte gemeinsam mit den Clusteranalyse-Algorithmen DBSCAN und OPTICS.

Grundprinzip

LOF definiert die "lokale Umgebung" eines Punktes über seine k nächsten Nachbarn. Der Abstand zu diesen wird verwendet, um eine lokale Dichte zu schätzen. In einem zweiten Schritt wird der Quotient aus den lokalen Dichten seiner Nachbarn und der lokalen Dichte des Punktes selbst gebildet. Dieser Wert bewegt sich nahe an 1.0 wenn ein Punkt in einem Bereich gleichmäßiger Dichte ist. Für Objekte, die aber abgeschieden von einer solchen Fläche sind, wird der Wert deutlich größer und kennzeichnet so Ausreißer.

Formal

Die k-Distanz(A) ist die Distanz des Objektes A zu seinem k-nächsten Nachbarn. Diese Menge kann gegebenenfalls mehr als k Objekte enthalten, wenn es mehrere Objekte mit dem gleichen Abstand gibt. Wir bezeichnen diese "k-Nachbarschaft" hier mit $N_k(A)$.



Kernidee von LOF: die lokale Dichte eines Punktes mit der seiner Nachbarn vergleichen.

Basierend auf dieser Distanz wird die "Erreichbarkeitsdistanz" definiert:

Erreichbarkeitsdistanz $_k(A, B) = \max\{k-\text{Distanz}(B), d(A, B)\}$ Die Erreichbarkeitsdistanz eines Objektes A von B ist also entweder der wahre Abstand, jedoch mindestens die k-Distanz von B. Objekte die zu den k-nächsten Nachbarn von B gehören werden also als gleich weit entfernt betrachtet. Die Motivation für diese Distanzdefinition ist es, stabilere Ergebnisse zu bekommen. (Es handelt sich hierbei aber nicht mehr um eine Distanzfunktion im mathematischen Sinne, da sie nicht symmetrisch ist.)

Die lokale Erreichbarkeitsdichte ("local reachability density", "lrd") eines Objektes A wird definiert als

$$\operatorname{lrd}(A) := 1 / \left(\frac{\sum_{B \in N_k(A)} \operatorname{Erreichbarkeits-Distanz}_k(A, B)}{|N_k(A)|} \right)$$

Diese Dichte ist also der Kehrwert der durchschnittlichen Erreichbarkeitsdistanz des Objektes *A von* seinen Nachbarn, nicht andersherum die

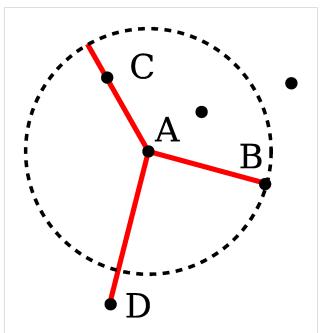


Illustration der Erreichbarkeitsdistanz. Objekte B und C haben die gleiche Erreichbarkeitsdistanz (k=3), während D kein k-nächster Nachbar ist

durchschnittliche Erreichbarkeitsdistanz der Nachbarn von A, was definitionsgemäß k-Distanz(A) wäre. Bei k+1 Duplikaten in einem Datensatz wird die Erreichbarkeitsdichte dieser Objekte unendlich.

Die lokale Erreichbarkeitsdichte wird jetzt mit denen der Nachbarn verglichen:

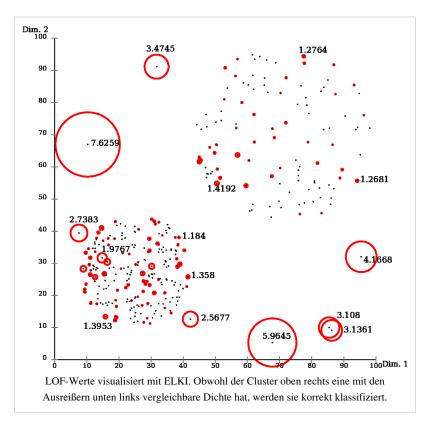
$$\mathrm{LOF}_k(A) := \frac{\sum_{B \in N_k(A)} \frac{\mathrm{lrd}(B)}{\mathrm{lrd}(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} \mathrm{lrd}(B)}{|N_k(A)|} / \mathrm{lrd}(A)$$

Der "Local Outlier Factor" (LOF) ist also die "Durchschnittliche Erreichbarkeitsdichte der Nachbarn" dividiert durch die Erreichbarkeitsdichte des Objektes selbst. Ein Wert von etwa 1bedeutet, dass das Objekt eine mit seinen Nachbarn vergleichbare Dichte hat (also kein Ausreißer ist). Ein Wert kleiner als 1bedeutet sogar eine dichtere Region (was ein sogenannter "Inlier" wäre), während signifikant höhere Werte als 1einen Ausreißer kennzeichnen.

Vorteile

Im Gegensatz zu vielen globalen Verfahren zur Ausreißererkennung LOF kann mit Bereichen unterschiedlicher Dichte in demselben Datensatz umgehen. Punkte mit einer "mittleren" Dichte in einer Umgebung mit "hoher" werden von LOF als Ausreißer klassifiziert, während ein Punkt mit "mittlerer" Dichte in einer "dünnen" Umgebung explizit nicht als solcher erkannt wird.

Während die geometrische Intuition von LOF nur in niedrigdimensionalen Vektorräumen Sinn ergibt, kann das Verfahren auf beliebigen Daten angewendet werden, auf denen eine "Unähnlichkeit" definiert werden kann. Es muss sich dabei nicht um eine Distanzfunktion im strengeren mathematischen Sinne handeln, die



Dreiecksungleichung wird beispielsweise nicht benötigt. Der Algorithmus wurde erfolgreich auf verschiedensten Datensätzen eingesetzt, beispielsweise zum Erkennen von Angriffen in Computernetzwerken, wo er bessere Erkennungsraten lieferte als die Vergleichsverfahren.

Nachteile

Ein wichtiger Nachteil von LOF ist, dass die Ergebniswerte schwer zu interpretieren sind. Werte um 1 und weniger sind sicher keine Ausreißer, aber es gibt keine klare Regel, ab welchem Wert ein Punkt ein *signifikanter* Ausreißer ist. In einem sehr gleichmäßigen Datensatz sind Werte von 1.1 auffällig, in einem Datensatz mit starken Dichteschwankungen kann ein Wert wie 2noch ein ganz normaler Datenpunkt sein. Im schlimmsten Falle, treten solche Unterschiede sogar in unterschiedlichen Teilen desselben Datensatzes auf.

Erweiterungen

- Feature Bagging for Outlier Detection wendet LOF in mehreren Projektionen an und kombiniert die Ergebnisse, um in hochdimensionalen Daten bessere Ergebnisse zu erhalten.
- Local Outlier Probability (LoOP) ist eine von LOF abgeleitete Methode, die die Dichte statistisch schätzt, um weniger abhängig vom genauen Wert von k zu werden. Zusätzlich wird das Ergebnis statistisch in den Wertebereich [0, 1] normalisiert, um besser interpretierbare Werte zu liefern.
- Interpreting and Unifying Outlier Scores stellt eine Normalisierung für LOF und andere Verfahren vor, die die Scores statistisch in das Intervall [0, 1] normalisiert um die Benutzerfreundlichkeit des Ergebnisses zu verbessern.

Verfügbarkeit

Eine Referenzimplementierung ist im Software-Paket ELKI verfügbar, inklusive Implementierungen von Vergleichsverfahren.

Einzelnachweise

Quelle(n) und Bearbeiter des/der Artikel(s)

Local Outlier Factor Quelle: http://de.wikipedia.org/w/index.php?oldid=122597752 Bearbeiter: Amygdala77, Andim, BuZZa, Chire, DampflokfanDR, HilberTraum, Kku, Käseliesel, Scooty, Simon Fischer, 7 anonyme Bearbeitungen

Quelle(n), Lizenz(en) und Autor(en) des Bildes

Datei:LOF-idea.svg Quelle: http://de.wikipedia.org/w/index.php?title=Datei:LOF-idea.svg Lizenz: Public Domain Bearbeiter: Chire

Datei:Reachability-distance.svg Quelle: http://de.wikipedia.org/w/index.php?title=Datei:Reachability-distance.svg Lizenz: Public Domain Bearbeiter: Chire

Datei:LOF.svg Quelle: http://de.wikipedia.org/w/index.php?title=Datei:LOF.svg Lizenz: Public Domain Bearbeiter: Chire

Lizenz

Wichtiger Hinweis zu den Lizenzen

THEOLOGY LITTINGES ZU UCH LAZEMZCH

Die nachfolgenden Lizenzen bezieht sich auf den Artikeltext. Im Artikel gezeigte Bilder und Grafiken können unter einer anderen Lizenz stehen sowie von Autoren erstellt worden sein, die nicht in der Autorenliste erscheinen. Durch eine noch vorhandene technische Einschränkung werden die Lizenzinformationen für Bilder und Grafiken daher nicht angezeigt. An der Behebung dieser Einschränkung wird gearbeitet.

Das PDF ist daher nur für den privaten Gebrauch bestimmt. Eine Weiterverbreitung kann eine Urheberrechtsverletzung bedeuten.

Creative Commons Attribution-ShareAlike 3.0 Unported - Deed
Diese "Commons Deed" ist lediglich eine vereinfachte Zusammenfassung des rechtsverbindlichen Lizenzvertrages (http://de.wikipedia.org/wiki/Wikipedia:Lizenzbestimmungen_Commons_Attribution-ShareAlike_3.0_Unported) ist allgemeinverständlicher Sprache.

Sie dürfen:

- das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen
 Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen
 Zu den folgenden Bedingungen:

- Namensnennung Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.
 Weitergabe unter gleichen Bedingungen Wenn Sie das lizenzierte Werk bzw. den lizenzierten Inhalt bearbeiten, abwandeln oder in anderer Weise erkennbar als Grundlage für eigenes Schaffen verwenden, dürfen Sie die daraufhin neu entstandenen Werke bzw. Inhalte nur unter Verwendung von Liteznzbedingungen weitergeben, die mit denen dieses Lizenzvertrages identisch, vergleichbar oder kompatibel sind.
- Verzichtserklärung Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten. Sonstige Rechte Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
- Die gesetzlichen Schranken des Urheberrechts und sonstigen Befugnisse zur privaten Nutzung;
 Das Urheberpersönlichkeitsrecht des Rechteinhabers;
 Rechte anderer Personen, entweder am Lizenzegegenstand selber oder bezüglich seiner Verwendung, zum Beispiel Persönlichkeitsrechte abgebildeter Personen
- Hinweis Im Falle einer Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für dieses Werk gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf http://creativecommons.org/licenses.by-sa/3.0/deed.de einzubinden.

Haftungsbeschränkung
Die "Commons Deed" ist kein Lizenzvertrag. Sie ist lediglich ein Referenztext, der den zugrundeliegenden Lizenzvertrag übersichtlich und in allgemeinverständlicher Sprache, aber auch stark vereinfacht wiedergibt. Die Deed selbst entfaltet keine juristische Wirkung und erscheint im eigentlichen Lizenzvertrag nicht.

GNU Free Documentation License

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.

51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA

Everyone is permitted to copy and distribute verbatim copies

of this license document, but changing it is not allowed.

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.
This License is a kind of "copyleft", which means that derivative works of the document must themselves essense. It complements the GNU General Public License, which is a copyleft license designed for free

software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or

1. APPLICABILITY AND DEFINITIONS

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document mans any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part at extbook of mathematics, a Secondary Section may not explain any mathematics). The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or back-Cover Texts, are certain secondary to the secondary of the part of the par

name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN OUANTITY

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover exts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full till ew with all words of the title equal by preminent and visible. You may add other manual on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

 B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

 C. State on the Title Page is authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

 C. State on the Title Page is authors, one or house of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

Lizenz 6

- D. Preserve all the copyright notices of the Document.
 E. Add an appropriate copyright notices for your modifications adjacent to the other copyright notices.
 F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
 G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

- F. Include, immediately after the copyright notices, a lacense notice giving the public permission to use the Modified Version under terms of this License, in the form shown in the Addendum below.

 G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

 H. Include an unaltered copy of this License.

 L. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, and publisher of the Document describing the Modified Version as stated in the previous sentence.

 J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for public access to a Transparent copy of the Document and likewise the network locations given in the Document for public access to a Transparent copy of the Document isself, or if the original publisher of the version it refers to give a publisher of the version at the version a

K. For any section Entitled Acknowledgements of Decument, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
 L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
 M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
 N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
 O. Preserve any Warranty Disclaimers.
 If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

as the authoritative definition of a standard.

You may not add a passage of up to five words as a Front-Cover Text, and a passage of up to 125 words as a Back-Cover Text, to the end of the list of Cover Text in the Modified Version. Only one passage of Front-Cover Text and a passage of up to five words as a Front-Cover Text and to expend the passage of the property of the passage of the property of the passage of the property of the passage of the pas

5. COMBINING DOCUMENTS

To unay combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of you combined work in sicense notice, and that you preserve all their warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple clauriant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedocsements".

6. COLLECTIONS OF DOCUMENTS

YOU MAY NOT DEFOCUTED IN 15
You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.
You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

1. AUGKEVATION WITH INDEFENDENT WOKES
A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

6. IACM SIZATION
Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the representation of the Invariant Section of this License or a notice or disclaimers, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

ed versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME.

Permission is granted to copy, distribute and/or modify this document

under the terms of the GNU Free Documentation License, Version 1.2

or any later version published by the Free Software Foundation;

with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled

"GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the

Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free