Notes on Neural Networks

Ivan Murashko

April 4, 2024

Contents

Introduction

TBD
[1]

Chapter 1

Neural network basics

1.1 Logistic regression

Lets consider a simple example of logistic regression that can be used for classification task. I.e. it will provide **yes** or **no** answer on a question about input data.

Suppose our input data are represented as the following vector

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

The result is a scalar $y \in \{1, 0\}$

In the model of supervised learning we have a set of m samples that are used for learning process:

$$\begin{gathered} \vec{x}^{(1)} \to y^{(1)}, \\ \vec{x}^{(2)} \to y^{(2)}, \\ \vdots \\ \vec{x}^{(m)} \to y^{(m)} \end{gathered}$$

We want to construct a function $\mathcal{Y}$ such that

$$\mathcal{Y}\left(\vec{x}^{(i)}\right) = y^{(i)}$$

for all $i \in \{1, \dots, m\}$. Taking into consideration that the result is a binary function, we can interpret $\mathcal{Y}$ as a probability that the outcome will be 1, i.e.

$$\mathcal{Y}\left(\vec{x}\right) = p\left(y = 1 | \, \vec{x}\right) \tag{1.1}$$

from eq. (1.1) we can derive

$$\begin{aligned} p\left(y=1 \middle|\, \vec{x}\right) &= \mathcal{Y}\left(\vec{x}\right), \\ p\left(y=0 \middle|\, \vec{x}\right) &= 1 - p\left(y=1 \middle|\, \vec{x}\right) = \\ &= 1 - \mathcal{Y}\left(\vec{x}\right) \end{aligned}$$

Thus we can conclude that

$$\begin{aligned} p\left(y \middle|\, \vec{x}\right) &= p\left(y=1 \middle|\, \vec{x}\right)^{y} \cdot p\left(y=0 \middle|\, \vec{x}\right)^{1-y} = \\ &= \mathcal{Y}\left(\vec{x}\right)^{y} \cdot \left(1 - \mathcal{Y}\left(\vec{x}\right)\right)^{1-y} \end{aligned} \tag{1.2}$$

For any test input $\vec{x}^{(i)} \to y^{(i)}$ we want to maximize the probability eq. (1.2). Taking into consideration that $p \geq 0$ and log is a monotonic function [3] then we can conclude that we want to maximize $\log p$ or

$$\log p = y \cdot \log\left(\mathcal{Y}\left(\vec{x}\right)\right) + (1-y)\log\left(1 - \mathcal{Y}\left(\vec{x}\right)\right) \tag{1.3}$$

Instead of maximizing $\log p$ eq. (1.3) we can minimize $-\log p$. Thus we will want to minimize the following cost function

$$\mathcal{L} = -y \cdot \log\left(\mathcal{Y}\left(\vec{x}\right)\right) - (1-y)\log\left(1 - \mathcal{Y}\left(\vec{x}\right)\right) \tag{1.4}$$

That is similar to Shannon entropy [2].

We want to minimize eq. (1.4) for each test input i.e.

$$\mathcal{L}^{(i)} = -y^{(i)} \cdot \log\left(\mathcal{Y}\left(\vec{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - \mathcal{Y}\left(\vec{x}^{(i)}\right)\right) \tag{1.5}$$

The eq. (1.5) can be combined into total cost function as follows

$$\mathcal{L} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \cdot \log\left(\mathcal{Y}\left(\vec{x}^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - \mathcal{Y}\left(\vec{x}^{(i)}\right)\right)\right] \tag{1.6}$$

we have not spoke about $\mathcal{Y}$ so far. Lets assume a linear dependency for it. it will consists of 2 functions: $\mathcal{Z}$ and σ. The first one is the $\mathcal{Z}$ that can be defined as follows

$$\mathcal{Z} = \vec{\omega}^{T}\vec{x} + b, \tag{1.7}$$

where

$$\vec{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}$$

i.e. the vector of the same dimension as $\vec{x}$. The equation eq. (1.7) can be rewritten as follows

$$\mathcal{Z} = \sum_{j=1} \omega_j x_j + b$$

and can be considered as a linear form on x_j. The eq. (1.7) cannot be interpret as a probability and we need a special function apply on it that is monotonic [3] and has its value between 0 and 1. The sigmoid function is our choice:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Combining all together we can write

$$\mathcal{Y}(\vec{\omega}, b, \vec{x}) = \sigma\left(\vec{\omega}^T \vec{x} + b\right) \tag{1.8}$$

The form eq. (1.8) is useful for computational purposes as soon as modern computer uses SIMD (single instruction multiple data) instraction set that allows to implement vector operations extremely efficient.

We have parameters $\omega_1, \ldots, \omega_n, b$ as unknown in eq. (1.8). We want to find them by finding a minimum for cost function eq. (1.6).

Contrary to eq. (1.8), our cost function is a function of unknown parameters only:

$$\begin{aligned} \mathcal{L}(\omega_1, \ldots, \omega_n, b) = -\frac{1}{m} \sum_{i=1}^{m} \Big[y^{(i)} \cdot \log\left(\mathcal{Y}\left(\omega_1, \ldots, \omega_n, b, \vec{x}^{(i)}\right)\right) + \\ + (1-y^{(i)}) \log\left(1-\mathcal{Y}\left(\omega_1, \ldots, \omega_n, b, \vec{x}^{(i)}\right)\right)\Big] \end{aligned} \tag{1.9}$$

For finding a minimum of $\mathcal{L}(\omega_1, \ldots, \omega_n, b)$ we can start with any point fig. 1.1, for instance $\omega_j = 0, \forall j$ and $b = 0$ as well. Then we should move in the direction opposite to gradient:

$$\begin{aligned} \omega_j^{(k)} &= \omega_j^{(k-1)} - \alpha \frac{\partial \mathcal{L}}{\partial \omega_j} \left(\omega_1^{(k-1)}, \cdots, \omega_n^{(k-1)}, b^{(k-1)}\right), \\ b^{(k)} &= b^{(k-1)} - \alpha \frac{\partial \mathcal{L}}{\partial b} \left(\omega_1^{(k-1)}, \cdots, \omega_n^{(k-1)}, b^{(k-1)}\right) \end{aligned}$$

where α is a parameter that relegates how fast we move to the minimum, k is the iteration counter.

Thus to find the final formula we have to calculate the derivatives. Lets

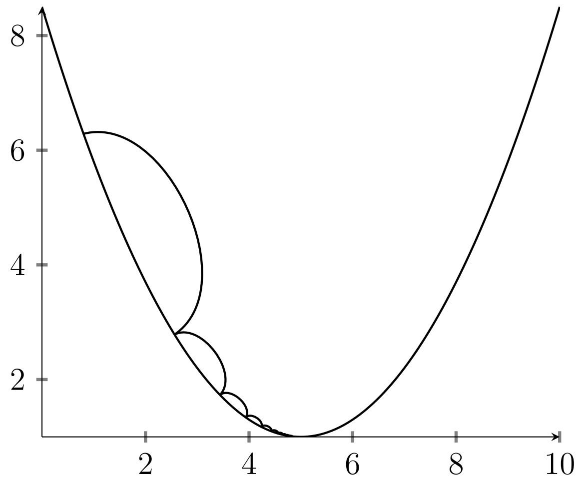

Figure 1.1: Gradient descent

rewrite our cost function eq. (1.9) for the easiest calculations

$$
\begin{aligned}
\mathcal{L}\left(\omega_1,\ldots,\omega_n,b\right) = \\
= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\cdot\log\left(\mathcal{Y}\left(\omega_1,\ldots,\omega_n,b,\vec{x}^{(i)}\right)\right)+\right. \\
\left.+\left(1-y^{(i)}\right)\log\left(1-\mathcal{Y}\left(\omega_1,\ldots,\omega_n,b,\vec{x}^{(i)}\right)\right)\right] = \\
= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left[\sigma\left(\sum_{j=1}^{n}\omega_j x_j^{(i)}+b\right)\right]+\right. \\
\left.+(1-y^{(i)})\log\left[1-\sigma\left(\sum_{j=1}^{n}\omega_j x_j^{(i)}+b\right)\right]\right]
\end{aligned}
$$

For instance

$$
\frac{\partial\mathcal{L}}{\partial\omega_j} = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\frac{1}{\sigma}\frac{d\sigma}{dz}\frac{\partial z}{\partial\omega_j}-(1-y^{(i)})\frac{1}{1-\sigma}\frac{d\sigma}{dz}\frac{\partial z}{\partial\omega_j}\right]
$$

where we used

$$
(\log(f))' = \frac{f'}{f}
$$

and used z as

$$
z = \sum_{j=1}^{n}\omega_j x_j^{(i)} + b
$$

i.e.
$$\frac{\partial z}{\partial \omega_j} = x_j^{(i)}$$
from the other size
$$\begin{aligned}\frac{d\sigma}{dz} &= -\frac{1}{(1+e^{-z})^2} \cdot \left(-e^{-z}\right) = \\ = \frac{1}{(1+e^{-z})^2}\left(1 - \frac{1}{1+e^{-z}}\right)\left(1+e^{-z}\right) &= \\ = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) &= \sigma(z)\left(1-\sigma(z)\right).\end{aligned}$$
Thus
$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \omega_j} &= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}x_j^{(i)}(1-\sigma(z)) - (1-y^{(i)})x_j^{(i)}\sigma(z)\right] = \\ &= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}x_j^{(i)} - \sigma(z)y^{(i)}x_j^{(i)} - x_j^{(i)}\sigma(z) + y^{(i)}x_j^{(i)}\sigma(z)\right] = \\ &= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}x_j^{(i)} - x_j^{(i)}\sigma(z)\right] = -\frac{1}{m}\sum_{i=1}^{m}x_j^{(i)}\left[y^{(i)} - \sigma(z)\right] = \\ &= \frac{1}{m}\sum_{i=1}^{m}x_j^{(i)}\left[\sigma(z) - y^{(i)}\right]\end{aligned}$$
where
$$\sigma(z) = \sigma\left(\vec{\omega}^T\vec{x}^{(i)} + b\right).$$
For derivative by b we have
$$\frac{\partial \sigma}{\partial b} = 1,$$
thus
$$\frac{\partial \mathcal{L}}{\partial b} = \frac{1}{m}\sum_{i=1}^{m}\left[\sigma(z) - y^{(i)}\right]$$

Bibliography

[1] Moggi, E. Notions of computation and monads. *Inf. Comput. 93*, 1 (1991), 55–92.

[2] Wikipedia contributors. Entropy (information theory) — Wikipedia, the free encyclopedia, 2024. [Online; accessed 3-April-2024].

[3] Wikipedia contributors. Monotonic function — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Monotonic_function&oldid=1208985525, 2024. [Online; accessed 4-April-2024].

Creative Commons Legal Code

Attribution-NonCommercial 3.0 Unported

> CREATIVE COMMONS CORPORATION IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS LICENSE DOES NOT CREATE AN ATTORNEY-CLIENT RELATIONSHIP. CREATIVE COMMONS PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. CREATIVE COMMONS MAKES NO WARRANTIES REGARDING THE INFORMATION PROVIDED, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM ITS USE.

License

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions

(a) **"Adaptation"** means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License.

(b) **"Collection"** means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.

(c) **"Distribute"** means to make available to the public the original and copies of the Work or Adaptation, as appropriate, through sale or other transfer of ownership.

(d) **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.

(e) **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.

(f) **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.

(g) **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

(h) **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.

(i) **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

2. Fair Dealing Rights. Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpet-

(a) Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

(b) Each time You Distribute or Publicly Perform an Adaptation, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.

(c) If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

(d) No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

(e) This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

(f) The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.

Creative Commons Notice