Семинар 11

Изкуствени невронни мрежи

В този семинар ще се запознаем с изкуствените невронни мрежи (ИНМ), разпространяващи сигналите напред (*forward feed*) и извършващи корекция на грешките в обратна посока (*back propagation of errors*). Те са изградени от няколко слоя неврони: входен слой, няколко скрити слоя и един изходен слой. На фигура 11.1 от лекция 11 е изобразена невронна мрежа с един скрит слой. Преобразуващата функция, с която ще се запознаем се дава с уравнение (11-I). Тази функция превръща чистия сигнал в числа между нула и единица.

$$Out_j = 1/[1 + \exp(-Net_j+offset_j)] \qquad \text{(11-I)}$$

Теоретичните въпроси са свързани с вида на тази функция. Подходът, който сме възприели е от стари учебници по математика, написани когато нямаше компютри или пък електронни калкулатори с графичен екран. Въпреки модерните графични програми, много често на изследователят му се налага да прецени вида и поведението на дадена функция в определен интервал и то в зависимост от параметрите във функцията[1]. Да определи дали функцията е винаги положителна или е отрицателна или сменя своя знак, дали е нарастваща или намаляваща, дали има локален максимум и/или локален минимум и дали те са глобалните максимум, съответно минимум в този интервал, или пък в крайните точки на интервала (наречени още гранични точки) са най-голямата или най-малката стойност на функцията.

Изложения по-долу подход се използва в много области на химичната наука и технология и затова запознаването с него е от полза за студентите.

Задача 11.1. Проверете какво става с функцията $y = 1/[1 + \exp(-x+const)]$ при клонене на x към -∞. А към +∞? Обърнете внимание, че това е

[1] Точно поради тези праметри едно изчертаване на функцията с компютърна програма не обобщава поведението на функцията.

преобразуващата функция от уравнение (11-I). Превърнете тази функция във вида y = exp(x-const)/[1 + exp(x-const)] като използвате равенството 1/exp(x) = exp(-x).

Решение: Вземаме предвид, че exp(x) е нарастваща функция[2]. Ако x е много малко, т.е. клони към -∞, то x-const клони също към -∞. Тогава -x+const клони към +∞, а exp(+∞) ще е много голямо число. Тогава знаменателят на функцията (11-I) ще е много голям и дробното отношение във функцията ще е много малко, т.е. функцията клони към нула.

Аналогично, ако x клони към +∞, то ще имаме exp(-∞) = 1/exp(+∞), което е нула. Тогава в знаменателя пренебрегваме експонента и дробното отношение става 1/1, т.е. единица. Което означава, че ако x клони към +∞ функцията (11-I) клони към единица.

Тук е мястото да се отбележи че функцията (11-I) е винаги положителна в целия интервал от -∞ до +∞. Причина за това е, че експонентът в знаменателя е винаги положителен.

За третото условие, преобразуването на функцията, ще разгледаме фукцията $y = 1/[1 + e^{-x}]$. Умножаваме числителя и знаменателя по e^x:

$$1/[1 + e^{-x}] = e^x/[e^x(1 + e^{-x})] = e^x/[e^x + e^x e^{-x})] = e^x/[e^x + 1]$$

Сега вместо x, заместваме x+const и получаваме и преобразувания вид на функцията (11-I): y = exp(x-const)/[1 + exp(x-const)].

Задача 11.2. Диференцирайте функцията от задача 11.1. Проверете че тя няма нито локален максимум, нито локален минимум и функцията е нарастваща в целия интервал от -∞ до +∞ т.е. първата производна е винаги положителна.

---

[2] За $a > 0$, може да се дефинира функцията $y = a^x$. Ако $a > 1$, то функцията е нарастваща, а при $a < 1$ - тя е намаляваща. Освен това, ако функцията y = f(x) е нарастваща, то функциите y = -f(x) и y = 1/f(x) са намаляващи.

Решение: Ще използваме няколко неща:

1. производната на e^x е пак e^x;

2. производната на $e^{z(x)}$ е $e^{z(x)}z'(x)$;

3. производната на $1/z(x) = [z(x)^{-1}]' = (-1)z(x)^{-2}z'(x)$.

Диференцираме функцията $y = 1/[1 + \exp(-x+const)]$. Получаваме:

$$y' = \{[1 + \exp(-x+const)]^{-1}\}' =$$

$$= (-1)[1 + \exp(-x+const)]^{-2}[1 + \exp(-x+const)]' =$$

$$= (-1)[1 + \exp(-x+const)]^{-2}\exp(-x+const)' =$$

$$= (-1)[1 + \exp(-x+const)]^{-2}\exp(-x+const)\mathbf{(-x+const)'} =$$

$$= (-1)[1 + \exp(-x+const)]^{-2}\exp(-x+const)\mathbf{(-1)} =$$

$$= \exp(-x+const)[1 + \exp(-x+const)]^{-2}$$

т.е.

$$y' = \exp(-x+const)[1 + \exp(-x+const)]^{-2} \qquad \text{(11-II)}$$

или написана, не като произведение, а като частно, производна е:

$$y' = \exp(-x+const)/\{[1 + \exp(-x+const)]^2\} \qquad \text{(11-IIa)}$$

Функцията $\exp(x)$ е винаги положителна, ето защо в числителя и знаменателя на (11-IIa) имаме положителни числа, т.е. първата производна е винаги положителна в интервала (-∞, +∞,), а това означава че функцията е нарастваща в целия интервал на реалните числа.

Интересно е да се види какво става с първата производна, когато аргументът x клони към -∞ или +∞. При -∞ експонентът е от големи положителни стойности и в знаменателя – вижте (11-IIa) - той е много по-голям от единицата. Тогава първата производна става

$$y' = \exp(-x+const)/[1 + \exp(-x+const)]^2 \approx$$

$$\approx \exp(-x+const)/[\exp(-x+const)]^2 =$$

$$= \exp(-x+const)^{-1} =$$

$$= 1/\exp(-x+const) = 1/\infty \approx 0$$

При +∞ експонентът е равен на много малка положителна стойност и в знаменателя – вижте (11-IIa) - той е много по-малък от единицата. Тогава първата производна става

$$y' = \exp(-x+const)/[1 + \exp(-x+const)]^2 \approx$$

$$\approx \exp(-x+const)/[1]^2 =$$

$$= \exp(-x+const) =$$

$$= \exp(-\infty+const) \approx 0$$

Това на практика означава, че функцията много бавно нараства в левия и десния край на интервала (-∞, +∞).

Задача 11.3. Получете втора производна на функцията от задача 11.1. и проверете, че тя е нула за $x = const$. В тази точка първата производна има максимална стойност, което ще рече, че преобразуващата функция се променя най-бързо в нея.

Решение: Освен свойствата на производните в предишната задача, ще използваме още няколко свойства:

1. Втора производна е производна на първата производна.

2. Производната на $z(x)w(x)$ е $z'(x)w(x) + z(x)w'(x)$;

$$y' = \exp(-x+const)[1 + \exp(-x+const)]^{-2} \qquad (11\text{-}II)$$

Изразът (11-II), за първата производна на функцията, е произведение на два множителя:

$$z(x) = \exp(-x+const) \quad \text{и} \quad w(x) = [1 + \exp(-x+const)]^{-2}$$

Производната на $z(x) = \exp(-x+const)$ е $\exp(-x+const)(-1)$, а производната на $w(x)$ е

$$w'(x) = (-2)[1 + \exp(-x+const)]^{-3}\,[1 + \exp(-x+const)]' =$$

$$= (-2)[1 + \exp(-x+const)]^{-3}\,[0 + \mathbf{\exp(-x+const)'}] =$$

$$= (-2)[1 + \exp(-x+const)]^{-3}\,\mathbf{\exp(-x+const)(-x+const)'} =$$

$$= (-2)[1 + \exp(-x+const)]^{-3} \exp(-x+const)(-1) =$$
$$= 2[1 + \exp(-x+const)]^{-3} \exp(-x+const)$$

Нека означим $-x+const$ с t за кратко изписване на формулите: после ще заместим $t = -x+const$. Тогава имаме

$$y''(x) = [y'(x)]' = [z(x)w(x)]' = z'(x)w(x) + z(x)w'(x) =$$
$$= -\exp(t)[1 + \exp(t)]^{-2} + \exp(t)2[1 + \exp(t)]^{-3} \exp(t)$$

общата част на двете събираеми по-горе е:

$$\exp(t)[1 + \exp(t)]^{-2}$$

Тогава изразът се преобразува в

$$y''(x) = \{\exp(t)[1 + \exp(t)]^{-2}\}\{-1 + 2\exp(t)[1 + \exp(t)]^{-1}\}$$

Първият множител (означен със синьо) е винаги положителен, въпреки че при малки t той клони към 0, а при големи t - също към нула.

Във втория множител можем да приведем към общ знаменател (който е $1 + \exp(t)$) и получаваме:

$$\{-1 + 2\exp(t)[1 + \exp(t)]^{-1}\} =$$
$$= \{-1 - \exp(t) + 2\exp(t)\}[1 + \exp(t)]^{-1} =$$
$$= [\exp(t) - 1][1 + \exp(t)]^{-1}$$

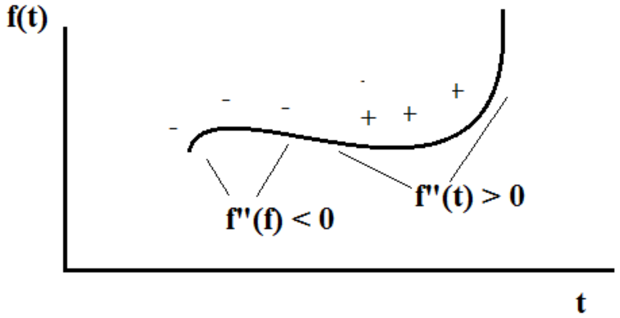
Накрая получихме

$$y''(x) = \{\exp(t)[1+\exp(t)]^{-2}\}[\exp(t) - 1][1+\exp(t)]^{-1} \qquad (11\text{-III})$$

Единственият множител (това е втория множител), който си сменя знака в (11-III) е $[\exp(t) - 1]$; останалите два множителя са по-големи от нула в целия интервал $(-\infty, +\infty)$. Ако приравним втората производна на нула, това означава че трябва да е изпълнено

$$[\exp(t) - 1] = 0$$

Това е изпълнено за $t = 0$, понеже $\exp(0) = 1$. Ако $t < 0$, имаме $[\exp(t) - 1] < 0$, а при $t > 0$ е изпълнено $[\exp(t) - 1]$. Отрицателна втора производна изначава, че функцията е вдлъбната, а положителна – че е изпъкнала. Това може да се запомни с изразите „събира дъжд“ и „не събира дъжд“ – вижте плюсовете и минусите на рисунката по-долу.



Припомнете си, че положихме $t = -x + const$. И така имаме:

$y''(x) < 0$ **при** $t < 0$, **което отговаря на** $x > const$

$y''(x) = 0$ **при** $t = 0$, **което отговаря на** $x = const$

$y''(x) > 0$ **при** $t > 0$, **което отговаря на** $x < const$

Задача 11.4. От резултатите от задачи 11.1 – 11.3 съобразете общия вид на функцията. Начертайте нейна приблизителна графика за x от -9 до + 11 при const = 1.0.

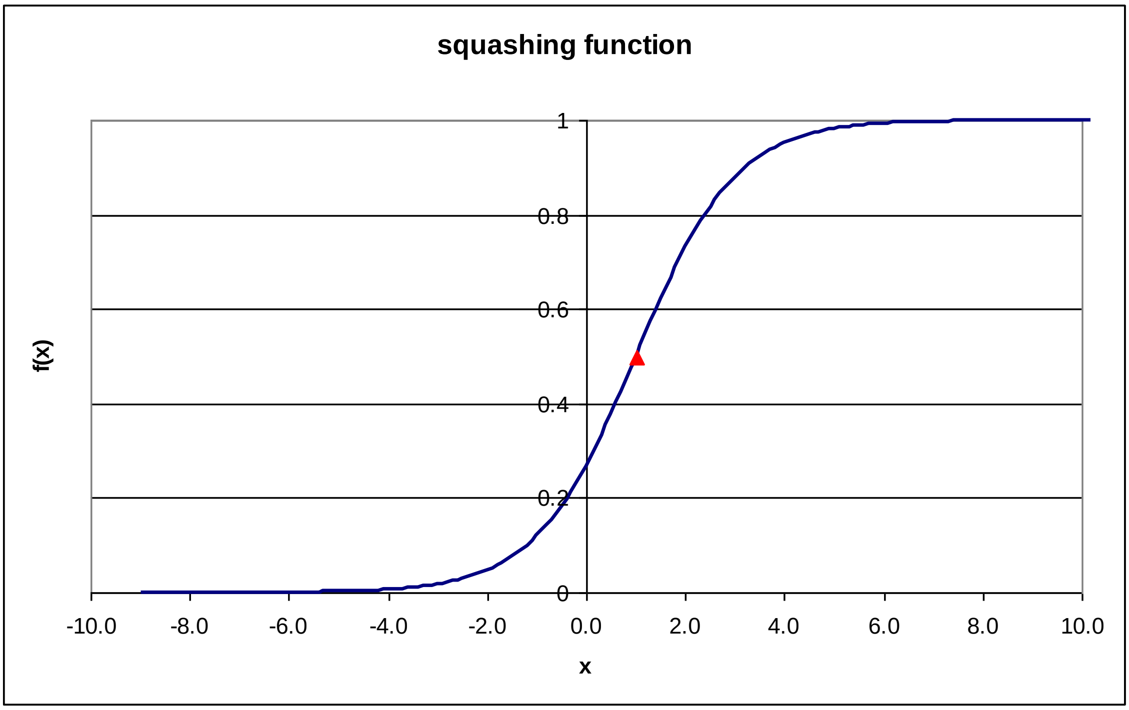
Решение: От задача 11.1 видяхме, че функцията (11-I) е винаги положителна в целия интервал от -∞ до +∞. В левия край на интервала (при -∞) тя клони към нула, но понеже е положителна се приближава отгоре към остта Ox. При много големи стойности на аргумента си, тя клони към единица, и очевидно ще се приближава до правата y = 1 от долния и край.

От задача 11.2 видяхме, че първата производна на функцията (11-I) е винаги положителна в целия интервал от -∞ до +∞. Това означава, че функцията е нарастваща. Видяхме, че в левия и десния край първата производна е много малка, т.е. там функцията нараства много бавно.

От задача 11.3 видяхме, че втората производна на функцията (11-I) е винаги положителна от -∞ до `const`, т.е. това е вдлъбната функция. При `x = const` втората производна е нула, което отговаря на инфлексна точка (сменя се кривината на функцията). В интервала от `const` до +∞ втората производна е отрицателна – функцията (11-I) изпъкнала в този интервал. Може да се докаже, че за `x = const` първата производна има максимум – т.е. в тази точка функцията нараства най-бързо. В тази точка функцията има стойност

```
y = 1/[1 + exp(-const + const)] = 1/[1+exp(0)] = 1/[1+1] = 0.5
```

Това позволява да се начертае на ръка графиката на функцията, както изисква задачата – за x от -9 до + 11 при const = 1.0. На следващата фигура е изчертана с Excel функцията в този интервал. С триъгълник на графиката е означена точката (`const, y(const)`).



Практически задачи

Задача C1. Отворете файла **`sem11_ann.xls`**. Разгледайте таблицата (sheet) "**`squashing function`**", в която са дадени изчислени стойностите на функцията (11-I). В клетка D1 може да променяте const (на английски се нарича offset). „Поиграйте си“ с няколко негови стойности. Какво става?

! Между всяка една функция y = f(x) и нейната родствена (изместена) функция y = f(x-const) съществува следното съответствие: при const > 0 графиката на втората функция е изместена надясно по остта Ox от първата функция с |const| - казва се, че функцията „закъснява“; при const < 0 графиката на втората функция е изместена наляво по остта Ox от първата функция с |const| - казва се, че функцията „избързва“.

Задача С2. Отворете файла sem11_ann.xls. Разгледайте таблицата (sheet) "primary alcohol", в която е приложен модел на ИНМ с два входни, два скрити и един изходен неврон. Спектралните признаци, които са вход в ИНМ се вкарват в региона B3:C3. Те се превръщат в проценти в региона B4:C4 и се автоскалират[3] в региона B8:C8. Съответни средни стойности и квадратите на стандартните отклонения (за обучаващата извадка) са в региона B5:C5 и, съответно, B6:C6. На практика, това което е в региона B8:C8 се нарича образ (pattern) и той се използва за работа с ИНМ. Коефициентите между входните и скритите неврони са дадени в региона E4:F5, между скритите и изходния неврон - в региона E9:F9. Съответните офсети са в региона E7:F7 и клетка E11. Преобразуващата функция при този модел е различна от тази в уравнение (11-I) – вместо да се изважда офсетът, той се събира: това нищо не променя принципно, освен няколко знака във формулите за изчисления[4].

В следната таблица са показани другите величини, които се изчисляват в ИНМ. Намерете клетките в таблицата на Excel и разгледайте форулите в тях. Вашата задача е да проследите във файла формулите, дадени в таблицата[5].

---

[3] За автоскалирането си припомнете от лекция 5!

[4] Тук става въпрос за преобразуване на функцията y = f(x) в y = f(x+const), вместо в y = f(x-const).

[5] Това е реално-действащ класификатор, който използва ИНМ и който е публикуван, наред с други за 19 химични фрагмента, в статията „P.N. Penchev, G.N. Andreev, K. Varmuza; *Automatic Classification of Infrared Spectra Using a Set of Improved Expert-based Features.* ***Anal. Chim. Acta,*** *388* (1-2), 145-159 (1999)“. Статията може да свалите от страницата на преподавателя.

Величина в ИНМ	Клетка в Excel файла	Формула за изчисление
$Net2_1$	B10	$Net2_1 = w21_{1,1}*x_1 + w21_{1,2}*x_2$
$Net2_2$	B11	$Net2_2 = w21_{2,1}*x_1 + w21_{2,2}*x_2$
$Out2_1$	B13	$Out2_1 = 1/[1+exp(-Net2_1-offset2_1)]$
$Out2_2$	B14	$Out2_2 =1/[1+exp(-Net2_2-offset2_2)]$
Net3 =	B16	$Net3 = w32_1*Out2_1 + w32_2*Out2_2$
Out3 =	B17	Out3 = 1/[1+exp(-Net3-offset3)]

Задача С3. В таблицата (sheet) "**`primary alcohol`**" на файла **`sem11_ann.xls`** проведете изчисления за спектрите на две от съединенията: Cyclohexanol, 2-methyl- и Cholesterol. Копирайте техните спектрални признаци, които се намират съответно в регионите **`K3:L3`** и **`K6:L6`**. Последователно залепете (paste) техните спектрални признаци в региона **`B3:C3`**: използвайте командата **`Paste Values`**. Какво става с изходът от ИНМ в клетка **`B17`**?

! Класификаторите на спектрална информация се разглеждат в дисциплината „Компютърни методи за обработка и интерпретация на спектрална информация", която се изучава в най-модерната и най-добрата магистратура на ПУ „П. Хилендарски", „Спектрохимичен анализ".

Задача С4. В таблицата (sheet) "**`primary alcohol`**" на файла **`sem11_ann.xls`** допълнително се изчисляват две вероятности - вероятност за отсъствие на структурния фрагмент (в този конкретен случай - първичен алкохол, CH_2OH) и вероятност за неговото присъствие. Използват се таблични данни, дадени в таблицата (sheet) "**`class limits`**". Вероятностите са дадени в клетки **`D21`** и **`D22`** и те са средното на вероятностите, които се получават като се използват стойностите в таблицата, които заграждат получения изход. Вашата задача е да разучите Excel функцията **`LOOKUP()`**. Отворете Help-a и прочете за тази функция, която е в раздела на **`Lookup and Reference`** функциите. Тук използваме нейния първи синтаксис.

В никой университет на Света знанията, които се придобиват на лекциите и семинарите, не са достатъчни. Хората стават добри специалисти като самостоятелно изучават книги, програми, помощни файлове и ръководства!