

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

9. Задача классификации

Постановка задачі

Задачи предсказания (напоминка)

Постановка задачи.

Есть множество объектов M с известными значениями признака Y .

Найти (предсказать, оценить) значение признака Y для нового объекта A .

Признак Y называется **целевым**.

Предсказываемый признак Y может быть:

- количественным;
- меткой класса.

В первом случае задача предсказания называется **задачей регрессии**, а во втором случае – **задачей классификации**.

Каждая из этих задач требует особой архитектуры нейронной сети.

Задачи классификации

Примеры.

- Предсказать пол человека по его фото (физическим данным).
- Предсказать наличие (отсутствие) болезни по результатам анализа.
- Предсказать результат командной многопользовательской игры по характеристикам и истории аккаунтов участников.
- Распознать тип дорожного знака для беспилотного автомобиля.
- По заданному тексту выяснить, сгенерирован он искусственным интеллектом или написан человеком.

Тренировочная выборка (ТВ)

В нашем курсе мы предполагаем, что все объекты для задачи предсказания описываются набором числовых признаков.

Иными словами, тренировочную выборку можно представить в виде таблицы. Для задачи классификации столбец Y содержит **метки класса**. Будем их обозначать с помощью натуральных чисел $\{0,1,2,\dots\}$.

Объект	Признак P_1	...	Признак P_n	Целевой признак Y
A_1	p_{11}	...	p_{1n}	y_1
A_2	p_{21}	...	p_{2n}	y_2
...
A_m	p_{m1}	...	p_{mn}	y_m

Пример ТВ для классификации

Важно помнить: не всё, что обозначается числами, является числом.

Метки класса – это не числа!

Поэтому решать задачу классификации как задачу регрессии плохо!

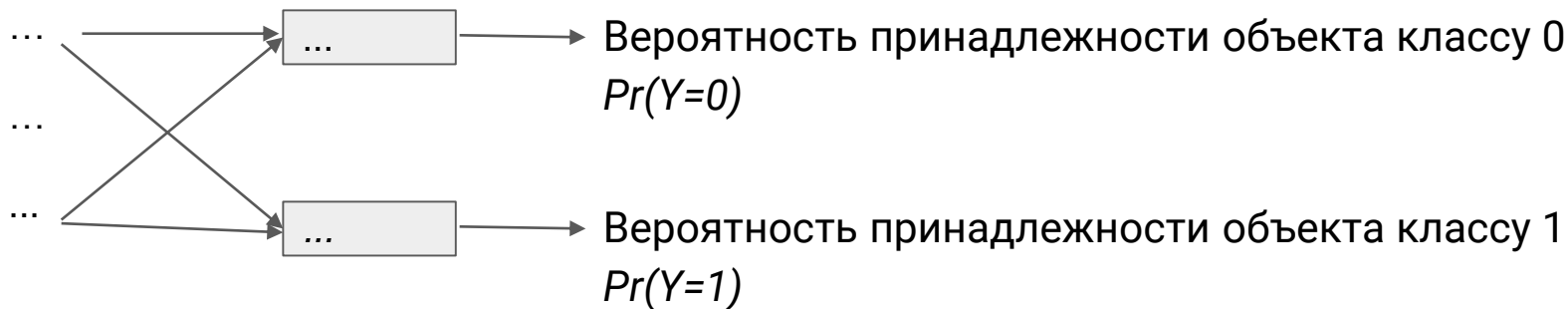
Пациент	Вес	Пол	Группа крови	Здоров? Y
A1	50	0	1	0
A2	60	1	2	1
A3	70	1	3	0
A4	80	0	4	1

Архитектура НС для классификации

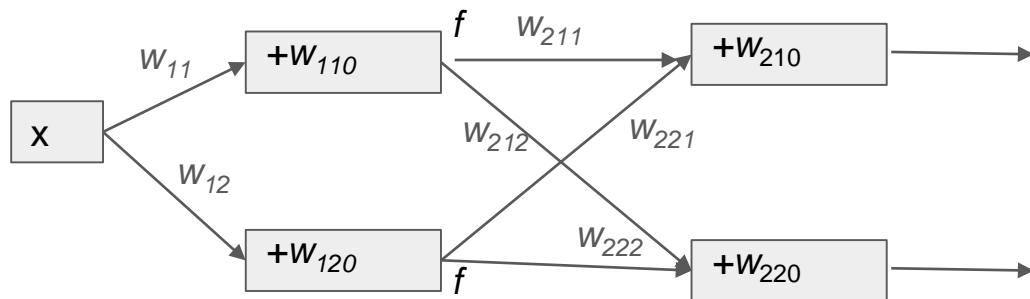
Как выглядит выходной слой НС, решающей задачу классификации?

Для простоты будем далее считать, что **классификация у нас бинарная**, то есть существует всего два (0,1) возможных значения целевого признака Y .

Тогда НС должна заканчиваться слоем из двух нейронов.



Самая простая сеть для классификации



Для объекта x выходы сети равны

$$R_0 = w_{211}f(xw_{11} + w_{110}) + w_{221}f(xw_{12} + w_{120}) + w_{210},$$

$$R_1 = w_{212}f(xw_{11} + w_{110}) + w_{222}f(xw_{12} + w_{120}) + w_{220}$$

Можно ли считать эти числа вероятностями принадлежности классам?

Есть проблемка...

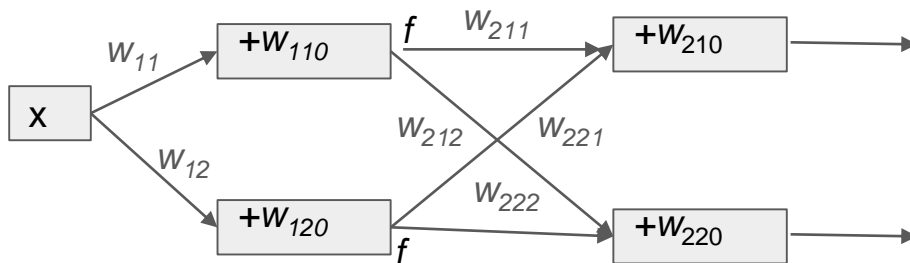
Выход сети и вероятности

Проблема в том, что числа, выдаваемые НС, могут не вполне годиться на роль вероятностей.

Они могут быть:

- не попадать в интервал $[0,1]$;
- их сумма может не быть равной 1.

Значит, нужно провести «умное» преобразование над числами, выдаваемыми НС.



Softmax

Пусть a, b - значения, которые выдал последний слой НС (не забывайте: у нас бинарная классификация).

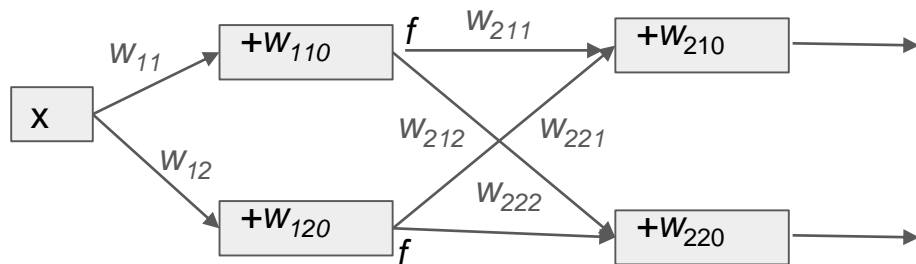
Тогда числа

$$\frac{e^a}{e^a + e^b}, \frac{e^b}{e^a + e^b}$$

положительны и их сумма равна 1. То есть их можно интерпретировать как вероятности принадлежности классам.

Операцию перехода от чисел a, b к указанным выше выражениям называется **операцией softmax**.

Самая простая сеть для классификации



Нужно к выходам сети R_1, R_2 применить softmax.

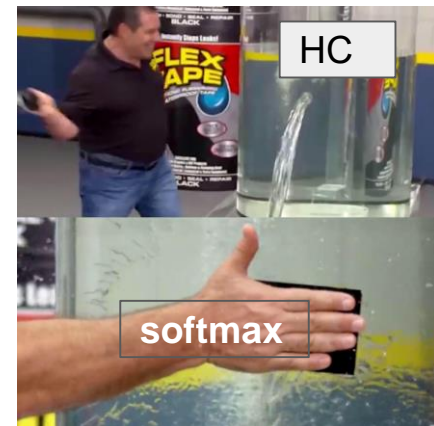
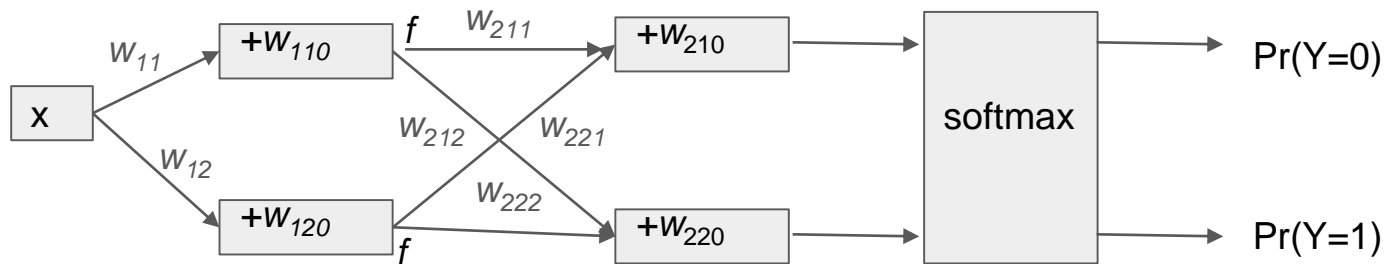
То есть сеть выдает вероятности

$$Pr(Y = 0) = \frac{e^{R_0}}{e^{R_0} + e^{R_1}}, \quad Pr(Y = 1) = \frac{e^{R_1}}{e^{R_0} + e^{R_1}}$$

$$R_0 = w_{211}f(xw_{11} + w_{110}) + w_{221}f(xw_{12} + w_{120}) + w_{210},$$
$$R_1 = w_{212}f(xw_{11} + w_{110}) + w_{222}f(xw_{12} + w_{120}) + w_{220}$$

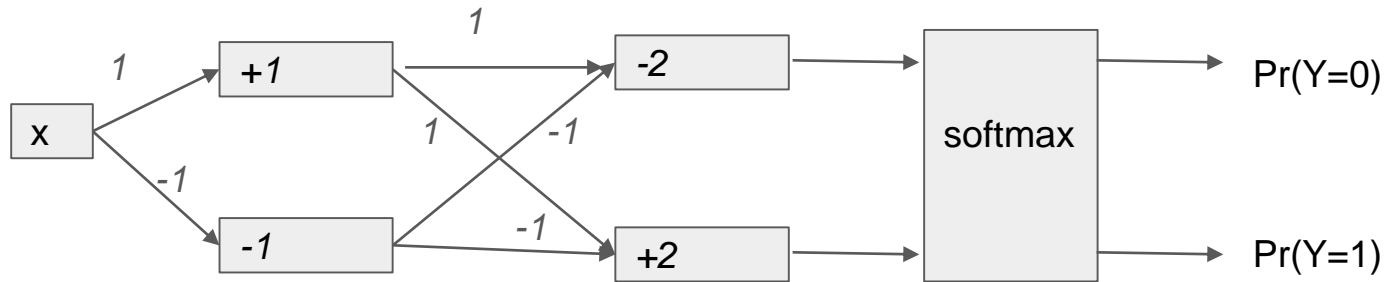
Слой softmax

Фактически преобразование softmax является дополнительным (и последним) слоем НС, решающей задачу классификации.



Получение вероятностей

Если веса НС уже натренированы, то для любого объекта можно получить вероятности принадлежности классам:



Например, если $x=0$, то на слой softmax приходят числа 0, 4. Softmax выдаст вероятности

$$Pr(Y=0)=1/(1+e^4)=0.02, Pr(Y=1)=e^4/(1+e^4)=0.98.$$

План решения задачи бинарной классификации

(полная аналогия с задачей регрессии)

1. **Взять тренировочную выборку**, то есть набор объектов с известными значениями целевого признака Y . Нейронная сеть должна восстановить зависимость между нецелевыми признаками и целевым признаком.
2. **Задать основные параметры нейронной сети**: количество слоёв, количество нейронов на каждом слое, тип связи между слоями и т.д.
3. **Выписать выражения для вероятностей $Pr(Y=0)$, $Pr(Y=1)$** . Эти выражения будут содержать вхождения весов w_i .
4. По ТВ составим функцию потерь $L(w)$ (**допустим, что мы уже умеем это делать**).

План решения задачи бинарной классификации

5. Функция потерь $L(w)$ содержит вхождения букв w_i (весов НС).
Относительно этих переменных мы **находим точку минимума функции $L(w)$** .
6. **Точка минимума определяет оптимальные веса НС.**
7. **Присваиваем весам НС найденные оптимальные значения.** Пусть теперь объект A не принадлежит тренировочной выборке. Мы прогоняем A через НС и на выходе получаем вероятности $Pr(Y=0)$, $Pr(Y=1)$ – они и являются предсказанием для объекта A .

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

9. Задача классификации

Функция потерь для классификации

Выбор функции потерь – это искусство

Очевидно, что нужно минимизировать выражение $\Pr(Y=0)$ для всех объектов ТВ из класса 1, и одновременно минимизировать выражение $\Pr(Y=1)$ для всех объектов ТВ из класса 0.

Или: **максимизировать вероятность принадлежности к истинному классу для каждого объекта из ТВ.**

Кросс-энтропия

Пусть ТВ состоит из объектов x_1, \dots, x_m , для которых известны их точные метки классов y_1, \dots, y_m , и НС даёт вероятности p_1, \dots, p_m принадлежности **к истинному классу**. Тогда в качестве функции потерь в задаче классификации берут выражение

$$L(w) = -\ln p_1 - \ln p_2 - \dots - \ln p_m$$

(**18+**: это **кросс-энтропия** распределений y_1, \dots, y_m и p_1, \dots, p_m)

Кросс-энтропия

$$L(w) = -\ln p_1 - \ln p_2 - \dots - \ln p_m$$

Насколько логично это выражение?

Во-первых, минимизация функции потерь $L(w)$ означает максимизацию выражения

$$L(w) = \ln p_1 + \ln p_2 + \dots + \ln p_m$$

Это означает, что нужно максимизировать аргументы логарифмов p_1, \dots, p_m .
Следовательно, нужно максимизировать вероятности принадлежности к истинному классу.

Пример идеального классификатора

Пусть для следующих объектов были получены такие вероятности принадлежности к их истинным классам:

Объекты	Истинная метка класса Y	Вероятность принадлежности к истинному классу
A1	0	1
A2	0	1
A3	1	1
A4	1	1

Тогда значение кросс-энтропии равно:

$L = -\ln(1) - \ln(1) - \ln(1) - \ln(1) = 0$ – и это **идеальный случай**.

Пример ужасного классификатора

Пусть для следующих объектов были получены такие вероятности принадлежности к их истинным классам:

Объекты	Истинная метка класса Y	Вероятность принадлежности к истинному классу
A1	0	0
A2	0	0
A3	1	0
A4	1	0

Тогда значение кросс-энтропии равно:

$L = -\ln(0) - \ln(0) - \ln(0) - \ln(0)$ = плюс бесконечность – и это **худший случай**.

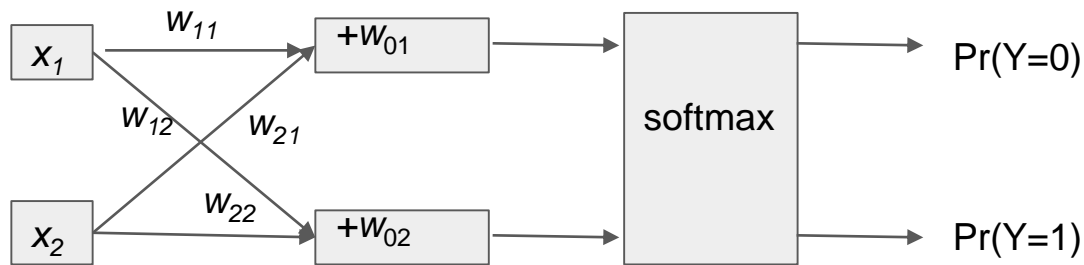
Значения кросс-энтропии на примерах

Объекты	Истинная метка класса Y	Вероятность принадлежности истинному классу
A1	0	0.9
A2	0	0.8
A3	1	0.8
A4	1	0.9
A5	1	0.7

$$L=-(\ln(0.9)+\ln(0.8)+\ln(0.8)+\ln(0.9)+\ln(0.7))=1.02$$

Пример ГС для классификации

Будем тренировать такую НС, классифицирующую объекты с двумя признаками x_1, x_2 (для простоты в ней нет ФА).



Выпишем выражения для выходов НС:

Пример ГС для классификации

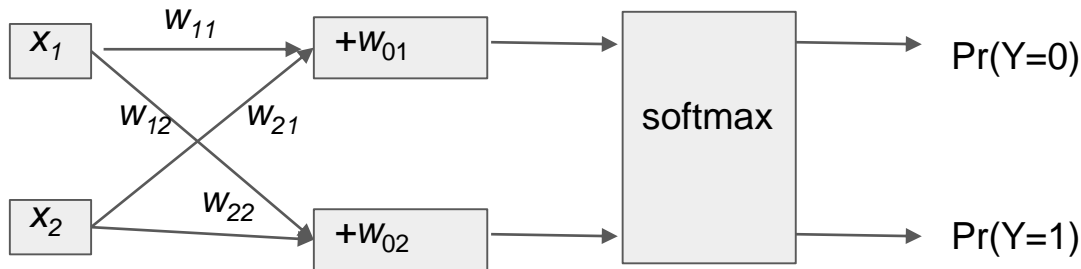
Выпишем выражения для выходов НС:

$$A = w_{11}x_1 + w_{21}x_2 + w_{01},$$

$$B = w_{12}x_1 + w_{22}x_2 + w_{02},$$

$$Pr(Y = 0) = \frac{e^A}{e^A + e^B},$$

$$Pr(Y = 1) = \frac{e^B}{e^A + e^B}$$



Пример ГС для классификации

Возьмем тренировочную выборку

Слагаемые функции потерь:

$$\ln \frac{e^{w_{11}+w_{01}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}},$$
$$\ln \frac{e^{w_{22}+w_{02}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$$

x_1	x_2	Y
1	0	0
0	1	1

$$A = w_{11}x_1 + w_{21}x_2 + w_{01},$$

$$B = w_{12}x_1 + w_{22}x_2 + w_{02},$$

$$Pr(Y = 0) = \frac{e^A}{e^A + e^B},$$

$$Pr(Y = 1) = \frac{e^B}{e^A + e^B}$$

$$L(w) = - \ln \frac{e^{w_{11}+w_{01}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}} - \ln \frac{e^{w_{22}+w_{02}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$$

$$L(w) = -\ln \frac{e^{w_{11}+w_{01}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}} - \ln \frac{e^{w_{22}+w_{02}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$$

$$\partial L / \partial w_{11} = \frac{-e^{w_{12}+w_{02}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}}$$

$$\partial L / \partial w_{22} = \frac{-e^{w_{21}+w_{01}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$$

$$\partial L / \partial w_{12} = 1 / (e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}})$$

$$\partial L / \partial w_{21} = 1 / (e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}})$$

$$\partial L / \partial w_{01} = \frac{-e^{w_{12}+w_{02}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}} + \frac{1}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$$

$$\partial L / \partial w_{02} = \frac{-e^{w_{21}+w_{01}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}} + \frac{1}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}}$$

В начале ГС все веса равны 0. Значения ЧП в этой точке:

$\frac{\partial L}{\partial w_{11}} = \frac{-e^{w_{12}+w_{02}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}}$	$\frac{\partial L}{\partial w_{11}}(a_0) = -0.5,$
$\frac{\partial L}{\partial w_{22}} = \frac{-e^{w_{21}+w_{01}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$	$\frac{\partial L}{\partial w_{22}}(a_0) = -0.5,$
$\frac{\partial L}{\partial w_{12}} = 1/(e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}})$	$\frac{\partial L}{\partial w_{12}}(a_0) = 0.5,$
$\frac{\partial L}{\partial w_{21}} = 1/(e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}})$	$\frac{\partial L}{\partial w_{21}}(a_0) = 0.5,$
	$\frac{\partial L}{\partial w_{01}}(a_0) = 0,$
	$\frac{\partial L}{\partial w_{02}}(a_0) = 0$
$\frac{\partial L}{\partial w_{01}} = \frac{-e^{w_{12}+w_{02}}}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}} + \frac{1}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}}$	
$\frac{\partial L}{\partial w_{02}} = \frac{-e^{w_{21}+w_{01}}}{e^{w_{21}+w_{01}} + e^{w_{22}+w_{02}}} + \frac{1}{e^{w_{11}+w_{01}} + e^{w_{12}+w_{02}}}$	

Тогда новые значения весов равны:

$$w_{11} := 0 - 0.1 * (-0.5) = 0.05,$$

$$w_{22} := 0 - 0.1 * (-0.5) = 0.05,$$

$$w_{12} := 0 - 0.1 * 0.5 = -0.05,$$

$$w_{21} := 0 - 0.1 * 0.5 = -0.05,$$

$$w_{01} := 0 - 0.1 * 0 = 0,$$

$$w_{02} := 0 - 0.1 * 0 = 0,$$

$$\partial L / \partial w_{11}(a_0) = -0.5,$$

$$\partial L / \partial w_{22}(a_0) = -0.5,$$

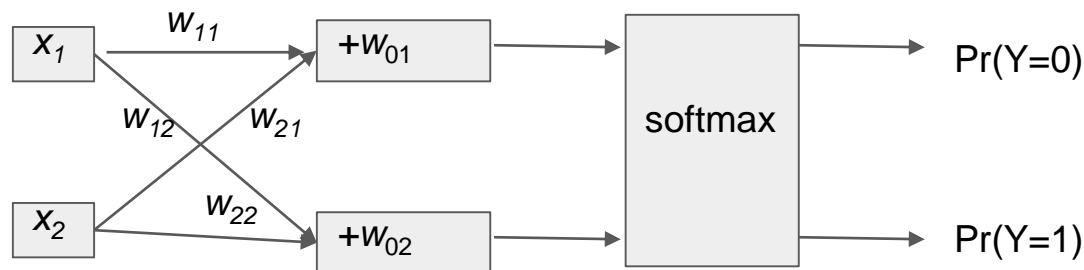
$$\partial L / \partial w_{12}(a_0) = 0.5,$$

$$\partial L / \partial w_{21}(a_0) = 0.5,$$

$$\partial L / \partial w_{01}(a_0) = 0,$$

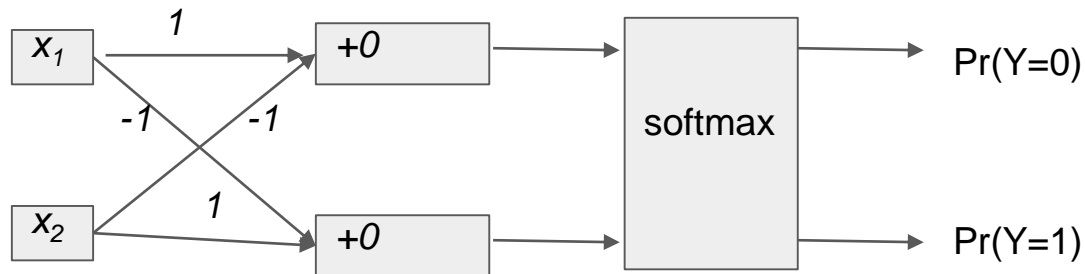
$$\partial L / \partial w_{02}(a_0) = 0$$

Насколько логично такое изменение весов для нашей НС и ТВ?



x_1	x_2	Y
1	0	0
0	1	1

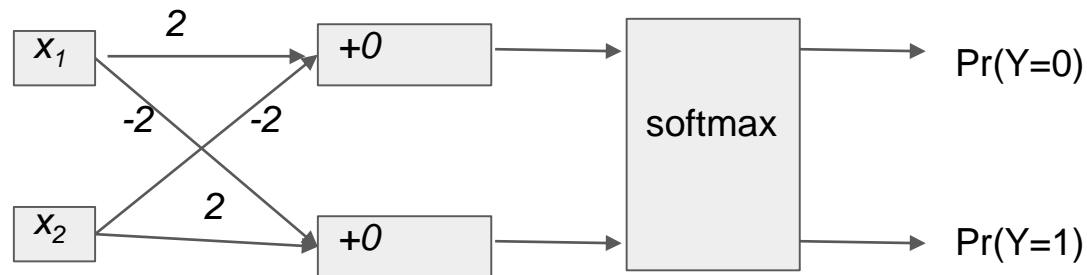
Через много итераций будет:



И наша сеть для объектов из ТВ предскажет вероятности:

x_1	x_2	Y	softmax	$\text{Pr}(Y=0)$	$\text{Pr}(Y=1)$
1	0	0	1, -1	0.88	0.12
0	1	1	-1, 1	0.12	0.88

А если еще подождать:



x_1	x_2	Y	softmax	$\text{Pr}(Y=0)$	$\text{Pr}(Y=1)$
1	0	0	2, -2	0.98	0.02
0	1	1	-2, 2	0.02	0.98

Хм... Это тот случай, когда ГС может длиться бесконечно.

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

9. Задача классификации

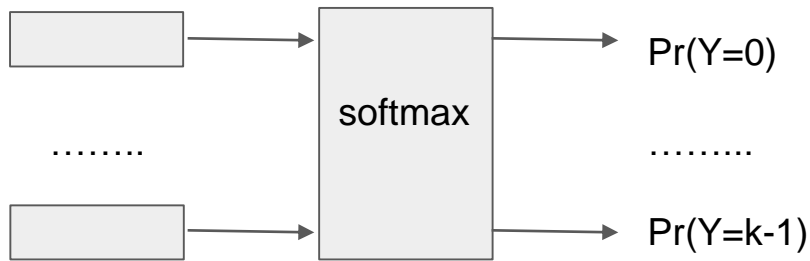
Многоклассовая классификация

Суть

До этого классификация была бинарной. А если классов больше, чем 2?

Всё аналогично!

Архитектура сети имеет **k выходов** (k – число классов).



В конце применяется **операция softmax**.

Многоклассовый softmax

Ожидаемая формула:

$$\begin{aligned} Pr(Y = 0) &= \frac{e^{a_0}}{e^{a_0} + \dots e^{a_{k-1}}}, \\ \dots \\ Pr(Y = k - 1) &= \frac{e^{a_{k-1}}}{e^{a_0} + \dots e^{a_{k-1}}} \end{aligned}$$

где a_0, \dots, a_{k-1} — выходы из нейронов последнего слоя.

Многоклассовая функция потерь

Для каждого объекта A_i ТВ можно вычислить величину p_i – вероятность принадлежности к **истинному классу**.

Как и раньше, функция потерь равна

$$L(w) = -\ln p_1 - \ln p_2 - \dots - \ln p_m$$

А дальше всё, как обычно: минимизируем эту функцию с помощью ГС.

Откуда вообще берётся логарифм в функции потерь?

Всё очень просто: мы же хотим, чтобы «вероятность p_i принадлежности истинному классу» у объекта A_i из ТВ была максимальной?

Значит, **нужно максимизировать все числа p_i** . То есть нужно максимизировать произведение $p_1 \cdot p_2 \cdot \dots \cdot p_m$

Максимизировать длинное произведение трудно, поэтому переходят к **максимизации логарифмов**: $\ln p_1 + \ln p_2 + \dots + \ln p_m$

Ну а максимизировать сумму логарифмов – это всё равно что умножить выражение на -1 и минимизировать. И мы получаем известную функцию потерь.

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

9. Задача классификации

Дурацкий классификатор

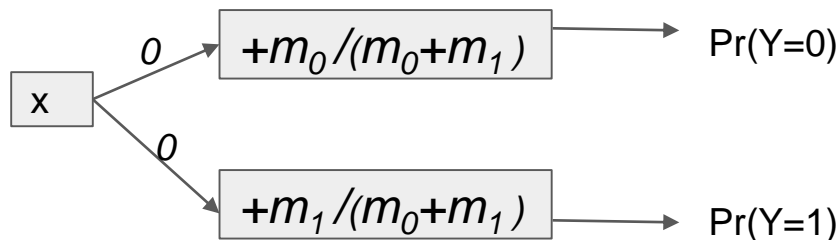


Дурацкий классификатор

Он работает так. Берёт ТВ: в ней m объектов, m_0 из них принадлежат классу 0 и m_1 принадлежат классу 1 ($m_0 + m_1 = m$). И **абсолютно для любого объекта классификатор выдает вероятности:**

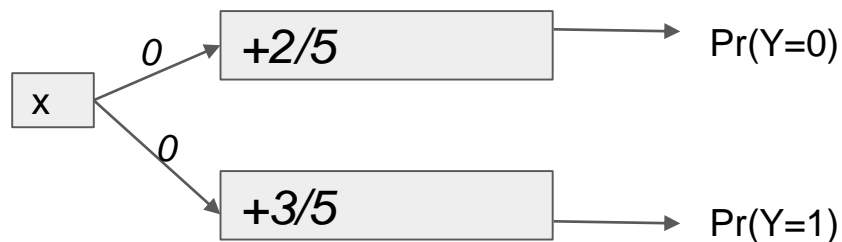
$$Pr(Y = 0) = \frac{m_0}{m_0 + m_1}, \quad Pr(Y = 1) = \frac{m_1}{m_0 + m_1},$$

Такой классификатор может быть реализован с помощью НС (даже без слоя softmax):



Тупой классификатор

Для этой ТВ тупой классификатор строит НС:



Объекты	Y
A1	0
A2	0
A3	1
A4	1
A5	1

то есть $\Pr(Y=0)=\frac{2}{5}$, $\Pr(Y=1)=\frac{3}{5}$ – у дурацкого классификатора вероятности не зависят от входа x (поэтому он и дурацкий).

Тупой классификатор

Объекты	Истинная метка класса Y	Вероятность принадлежности истинному классу
A1	0	0.4
A2	0	0.4
A3	1	0.6
A4	1	0.6
A5	1	0.6

$$L = -(\ln(0.4) + \ln(0.4) + \ln(0.6) + \ln(0.6) + \ln(0.6)) = 3.36$$

У вашей НС функция потерь должна быть ниже, чем у тупого классификатора.

Значение функции потерь

Можно вычислить значение функции потерь дурацкого классификатора

$$-m_0 \ln \frac{m_0}{m_0 + m_1} - m_1 \ln \frac{m_1}{m_0 + m_1}$$

Эта величина позволяет грубо оценить силу вашей НС.

Задача

Пусть ТВ состоит из 10 объектов: 6 объектов принадлежат классу 0, 4 объекта принадлежат классу 1. Дата-саентист Вася натренировал НС такую, что значение функции потерь на ТВ равно 6.75.

Задание: огорчите Васю.

Решение: НС Васи можно сравнить с дурацким классификатором, который для данной ТВ имеет следующее значение функции потерь:

$$-m_0 \ln \frac{m_0}{m_0 + m_1} - m_1 \ln \frac{m_1}{m_0 + m_1} = -6 \ln 0.6 - 4 \ln 0.4 = 6.74$$

То есть дурацкий классификатор сильнее, чем НС Васи.

КЛАССИФИКАТОРАМИ

С такими ~~приколами~~ тебе сюда, Вася



Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

9. Задача классификации

Выводы

Выводы:

- Мы рассмотрели постановку задачи классификации.
- Была рассмотрена специальная архитектура НС, необходимая для решения задачи классификации.
- Была изучена операция softmax, преобразующая выходы НС в вероятности.
- Для задачи классификации была найдена специальная функция потерь (кросс-энтропия).
- Были найдены оценки точности НС для задачи классификации (точность не должна быть ниже точности «дурацкого» классификатора).