

Методи машинного навчання для класифікації мобільного абонента як водія

Виконала: студент 4 курсу, групи КА-31

Мулява Галина Ярославівна

Керівник асистент Кухарев С.О.

Мета дослідження:

1. Дослідження існуючих методів машинного навчання, зокрема штучних нейронних мереж
2. Розробка програмного продукту, яке дає змогу класифікувати мобільного абонента за його характеристиками як автомобіліста

Об'єкт і предмет:

- Об'єктом дослідження є ознаки , зібрані з абонентської бази ПрАТ Київстар
- Предметом дослідження є застосування нейронних мереж для задачі класифікації

Основний результат:

Реалізація програмного продукту, який з високою точністю розв'язує задачу класифікації

Формальна постановка задачі

X – множина об'єктів, Y – множина допустимих відповідей

$y^*: X \rightarrow Y$ цільова функція

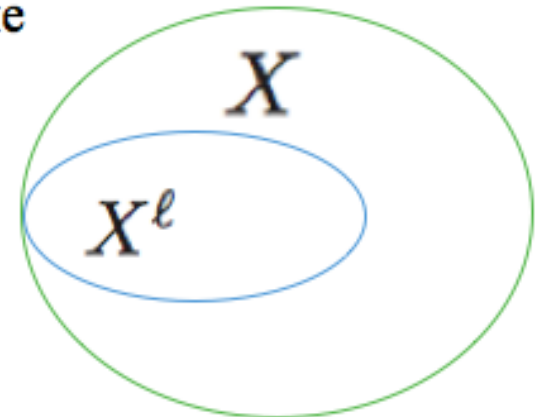
$y_i = y^*(x_i)$ значення відомі на $X^\ell = (x_i, y_i)_{i=1}^\ell$

Задача навчання

за вибіркою X^ℓ відновити залежність y^*

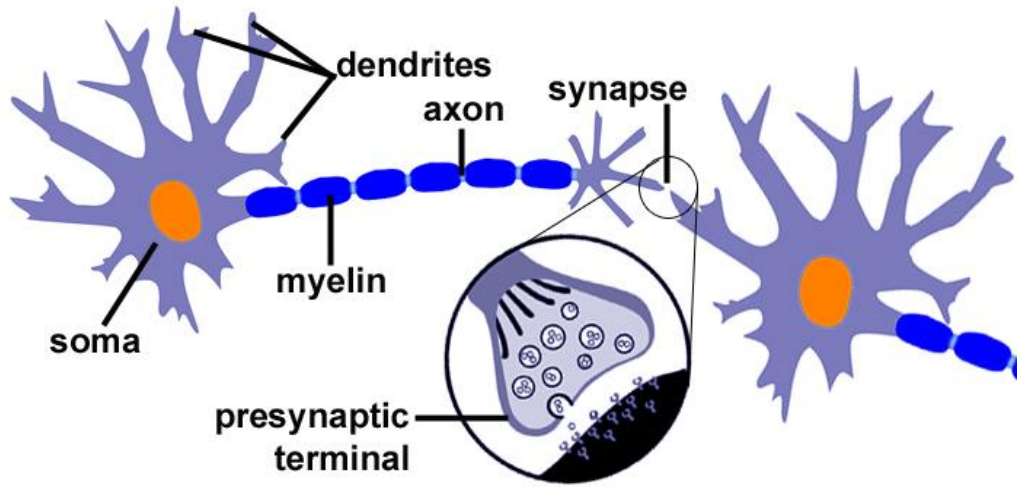
Тобто побудувати вирішальну функцію:

$a: X \rightarrow Y$ яка наближала б цільову функцію, причому не тільки на об'єктах навчальної вибірки, а й на всій множині X .

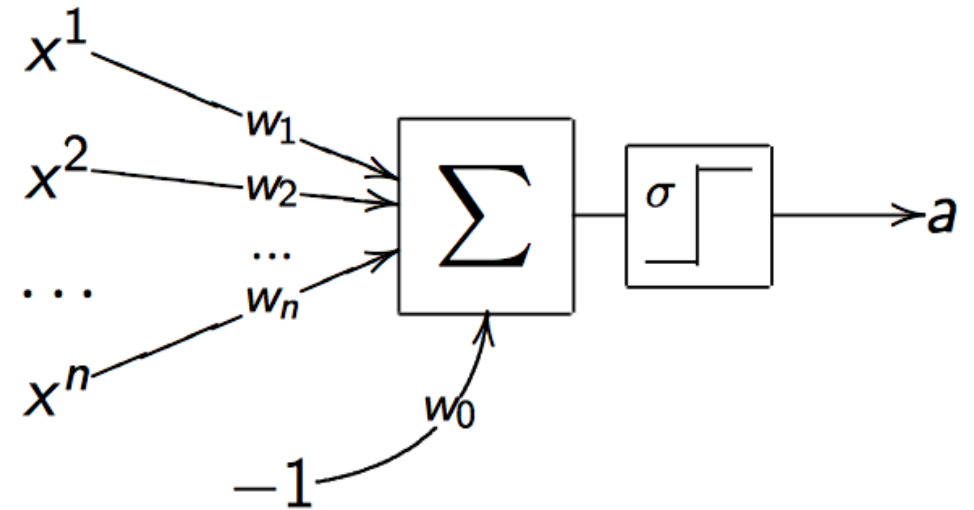


Нейрон

Біологічний



Штучний



$w_0, w_1, \dots, w_n \in \mathbb{R}$ ваги ознак

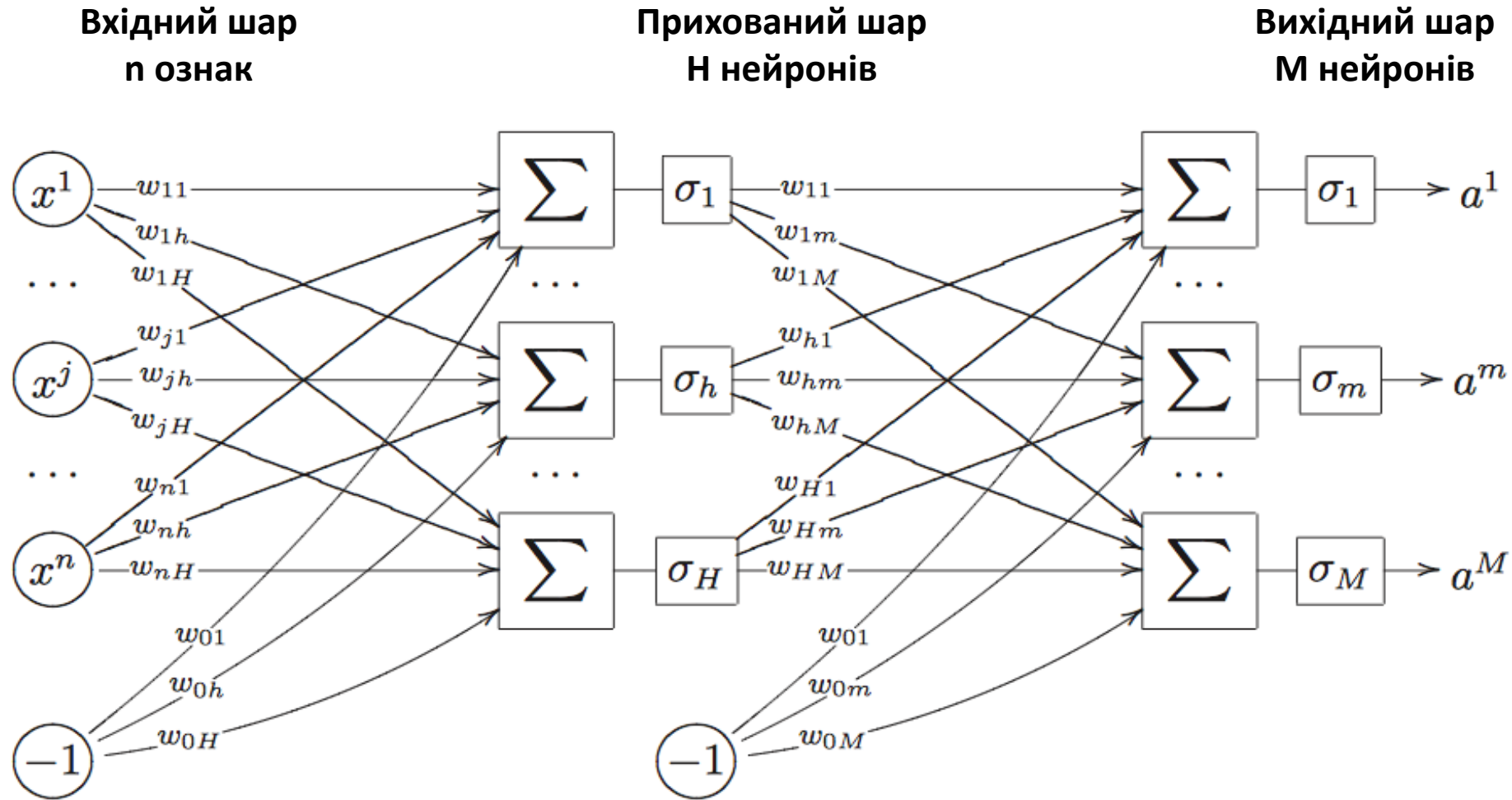
$\sigma(s)$ функція активації (нап. sign)

f_j ознака об'єкта x , $j = 1, \dots, n$

$$a(x, w) = \sigma(\langle w, x \rangle) = \sigma\left(\sum_{j=1}^n w_j f_j(x) - w_0\right)$$

Нейронна мережа

- Нейронна мережа = суперпозиція нейронів з нелінійною функцією активації



Вибір параметрів моделі

Хочемо знайти:

$$w \equiv (w_{jh}, w_{hm}) \in \mathbb{R}^{H(n+M+1)+M}$$

Отримуємо задачу оптимізації:

$$Q(w) := \sum_{i=1}^{\ell} \mathcal{L}(w, x_i, y_i) \rightarrow \min_w$$

Функція втрат

$$\mathcal{L}(a, x) = [a(x) \neq y(x)]$$

$$\mathcal{L}(a, x) = \frac{1}{(1 + \exp(-ax))}$$

Метод стохастичного градієнтного спуску
та

Розв'язуємо

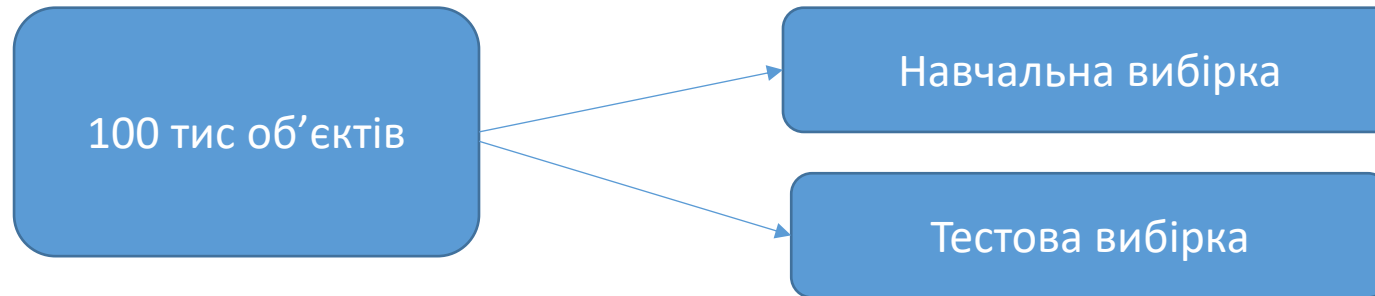
Метод зворотного поширення помилки

Вхідні дані

- Об'єкт: абонент
- Ознаки об'єкту:

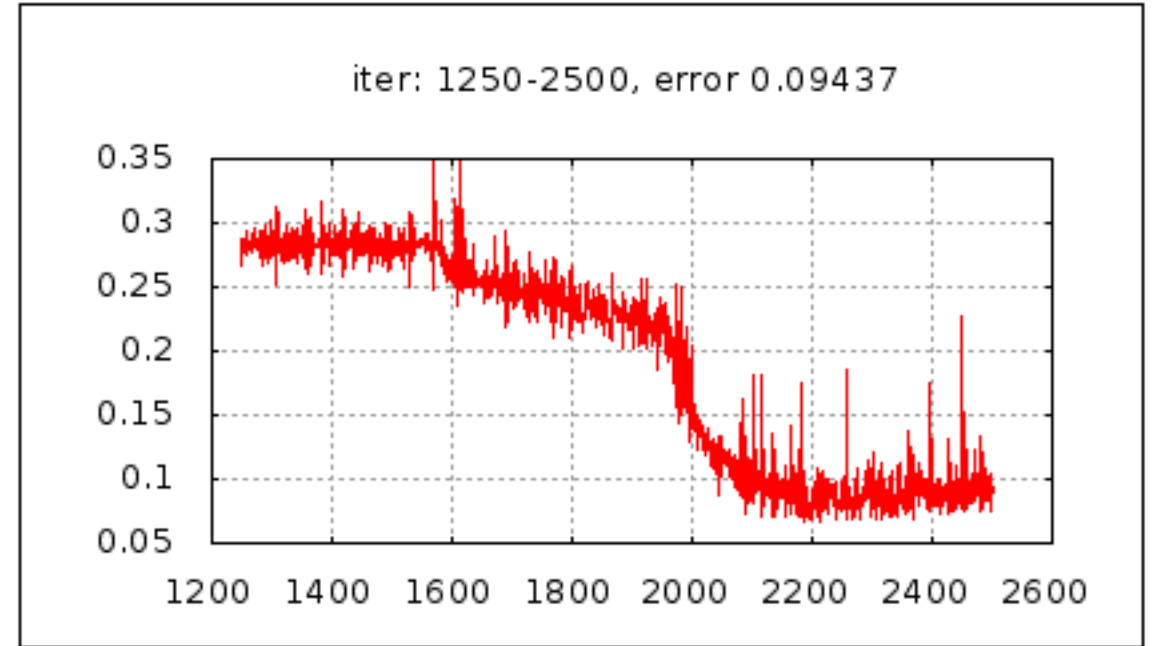
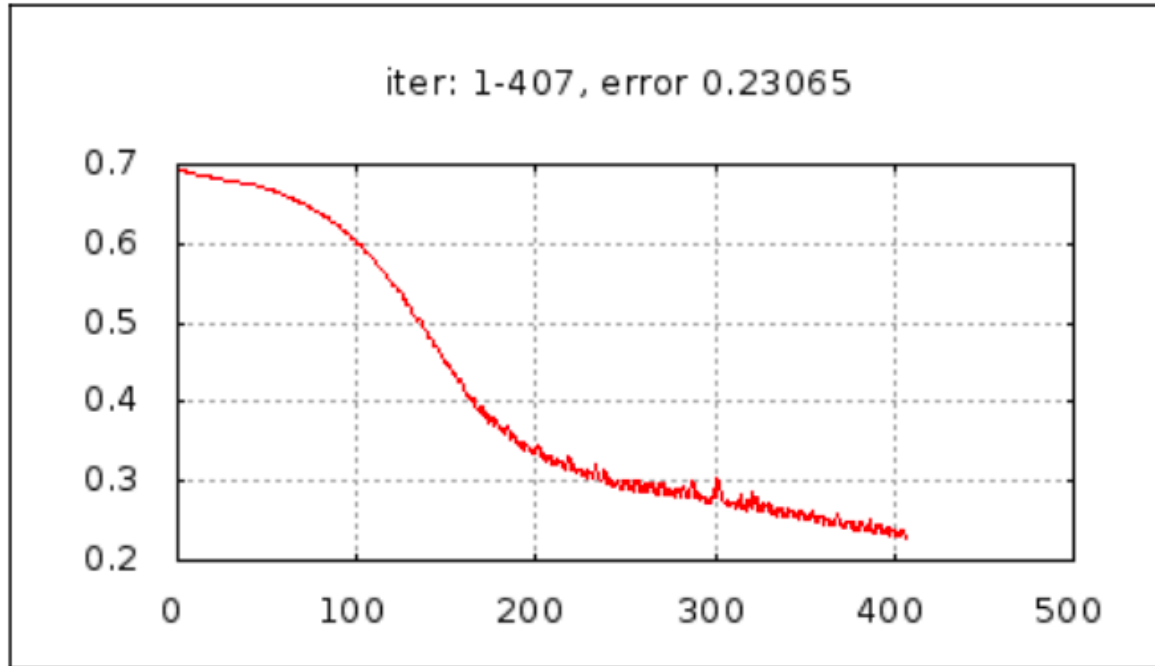
Помісячно середня	К-ть повідомлень	Заправки / автосервіси / магазини деталей / шиномонтаж / автомийка
	К-ть дзвінків	
	К-ть вебсайтів	
	К-ть базових станцій мобільного зв'язку	
Єдине число	Стать	Характеристики людини
	К-ть років	
	Передоплатний зв'язок/ Контракт	

Навчання мережі



Функція активації	Точність
$f(x) = x$	76%
Логістична $f(x) = 1 / (1 + \exp(-x))$	92%
$f(x) = \max(0, x)$	91%

Результати

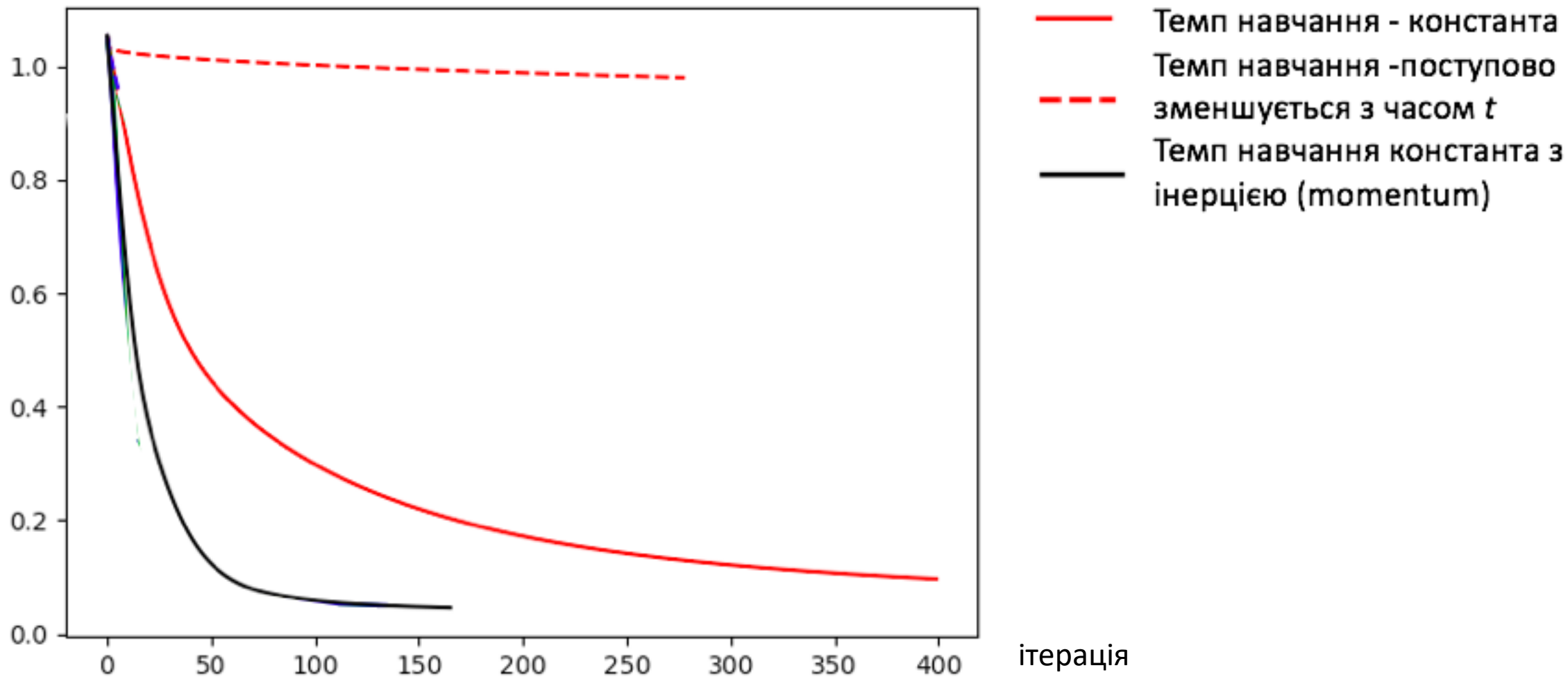


Метод стохастичного градієнтного спуску

Нейронна мережа з трьома шарами: вхідний шар 7 нейронів, перший прихований шар 100 нейронів з активацією Relu, другий шар - 2 нейрона (за числом класів) з активацією Log. Мережа досягла порогу помилки за 2543 епохи.

Оцінка роботи методу стохастичного градієнтного спуску

Функція помилки



Оцінка роботи

Метод кросвалідації (англ. cross-validation)

	Багатошарова перцептронна нейронна мережа		
Кількість прихованих шарів	1	3	10
Кількість нейронів	100	50	15
	Результат		
Складність алгоритму	10^3	$3 \cdot 10^3$	$2 \cdot 10^3$
Точність	89 %	95 %	91%

Висновки

- Було реалізовано багатошарову перцептронну мережу прямого поширення для задачі класифікації мобільних абонентів
- Визначено функцію втрат, оптимальна кількість прихованих шарів та їх кількість нейронів, функції активації
- Було обрано метод стохастичного градієнтного спуску і його параметри
- Обрано параметри регуляризації моделі
- Найкраща точність становить 95%

Перспективи подальших досліджень

- Відбір ознак (*feature selection*) (збільшимо кількість характеристик об'єкта і виберемо найсуттєвіші)
- Задача кластеризації всередині класів за допомогою інших методів машинного навчання неконтрольованого (« без учителя»)

Додається акт впровадження результатів дослідження у ТОВ «РДТЕХ»

Дякую за увагу!