

Разделение по вариантам:

1 Вариант:	2 Вариант:
Уткина	Апанасенко
Беденко	Якунина
Букина	Беликова
Горохова	Гаврищук
Карапетьян	Крылова
Манузина	Лян Чи Цзюн
Никитин	Путнин
Плешакова	Рунович
Рондону	Фейгина

Контрольная работа №2

1. Используя таблицу ASCII и таблицу перевода чисел из одной системы счисления в другую: представьте следующие выражения в двоичной системе счисления:

1 ВАРИАНТ: «Hello, World!», «short-term», « $12 + x = y - 4:z$ » (пробел тоже считается за знак)

2 ВАРИАНТ: «Good luck!», «long-term», « $10.5 + 1.3 \geq y - 5 \cdot x$ » (пробел тоже считается за знак)

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

Десятичное число	Восьмеричное число	Двоичная запись	Шестнадцатеричное число	Двоичная запись
0	0	000	0	0000
1	1	001	1	0001
2	2	010	2	0010
3	3	011	3	0011
4	4	100	4	0100
5	5	101	5	0101
6	6	110	6	0110
7	7	111	7	0111
8	10	000	8	1000
9	11	001	9	1001
10	12	010	A	1010
11	13	011	B	1011
12	14	100	C	1100
13	15	101	D	1101
14	16	110	E	1110
15	17	111	F	1111

2.

Реализуйте one hot encoding и n-gram модель($n = 2$) для слов из следующего предложения:

1 ВАРИАНТ: «There are beautiful flowers growing in the garden»

2 ВАРИАНТ: «Read a book about the history of America.»

Ответ представьте в следующем виде:

one hot encoding: word = [1,0,0...]

n-gram: (w_1, w_2), (w_2, w_3) ... (w_n, w_m), где w_i - соответствующее слово

3.

1 ВАРИАНТ: Допустим, документ содержит 100 слов, в котором слова «кошка» встречается 5 раз. Чему равен term-frequency (tf) для данного слова? Теперь предположим, что у нас есть 10 миллионов текстов и слово «кошка» встречается в тысячи из них. Чему равна inverse document frequency (idf)? Чему в итоге равен tf-idf вес для слова «Кошка»?

2 ВАРИАНТ: Допустим, документ содержит 100 слов, в котором слова «человек» встречается 2 раза. Чему равен term-frequency (tf) для данного слова? Теперь предположим, что у нас есть 1,5 тысячи текстов и слово «человек» встречается в ста из них. Чему равна inverse document frequency (idf)? Чему в итоге равен tf-idf вес для слова «человек»?

4.

Какие архитектуры нейронных сетей для перевода слов в вектора вы знаете, в чем их принципиальные отличия? Приведите пример и нарисуйте базовую схему сети. (не более 2)

1 ВАРИАНТ: Допустим имеется модель CBOW (Continuous Bag of Words). На вход сети подается 2 закодированных с помощью one-hot encoding слова из нижеприведенного текста. То есть на вход сети подаются 2 вектора размерности V . Определить количество нейронов на входном и выходном слоях. Какова будет размерность матрицы весов между входным и скрытым слоями, если скрытый слой имеет 10 нейронов?

Текст: «*Maître Corbeau, sur un arbre perché, Tenait en son bec un fromage. Maître Renard, par l'odeur alléché, Lui tint à peu près ce langage : « Hé ! bonjour, Monsieur du Corbeau. »»*

Подсказка: Необходимо найти V с учетом вышеприведенного текста.

2 ВАРИАНТ: Допустим имеется модель Skip-Gram. На вход сети подается закодированный с помощью one-hot encoding вектор размерности V . Слово, которому соответствует данный вектор определяется из нижеприведенного текста. Определить количество нейронов на входном и выходном слоях нейронной сети. Какова будет размерность матрицы весов между входным и скрытым слоями, если скрытый слой имеет 5 нейронов?

Текст: «*Jeder hat das Recht auf Bildung. Die Bildung ist unentgeltlich, zum mindesten der Grundschulunterricht und die grundlegende Bildung. Der Grundschulunterricht ist obligatorisch.*»

Подсказка: Необходимо найти V с учетом вышеприведенного текста.

5.

В чем принципиальное отличие архитектуры LSTM от стандартной RNN ?

1 ВАРИАНТ: Изобразите модуль (ячейку) обыкновенной рекуррентной сети. В чем состоит отличие/отличия архитектуры RNN от стандартной архитектуры нейронной сети? Каковы недостатки стандартной рекуррентной сети?

2 ВАРИАНТ: Изобразите модуль (ячейку) сети LSTM. Пошагово объясните каждый процесс в ячейке. В чем состоит основная идея LSTM?

6.

Что такое Doc2vec? Какие архитектуры нейронных сетей являются наиболее популярными doc2vec? в чем их состоит их сходство с соответствующими моделями из word2vec?