Machine Learning Exercise Sheet 07

Deep Learning

In-class Exercises

See the recording of the in-class exercise for the discussion of the code in the notebooks.

Problem 1: See notebook exercise_inclass_07_vectorization_numerics.ipynb on Moodle.

Problem 2: See notebook exercise_inclass_07_backpropagation.ipynb on Moodle.

Homework

Problem 3: In machine learning you often come across problems which contain the following quantity

$$y = \log \sum_{i=1}^{N} e^{x_i}$$

For example if we want to calculate the log-likelihood of neural network with a softmax output we get this quantity due to the normalization constant. If you try to calculate it naively, you will quickly encounter underflows or overflows, depending on the scale of x_i . Despite working in log-space, the limited precision of computers is not enough and the result will be ∞ or $-\infty$.

To combat this issue we typically use the following identity:

$$y = \log \sum_{i=1}^{N} e^{x_i} = a + \log \sum_{i=1}^{N} e^{x_i - a}$$

for an arbitrary a. This means, you can shift the center of the exponential sum. A typical value is setting a to the maximum ($a = \max_i x_i$), which forces the greatest value to be zero and even if the other values would underflow, you get a reasonable result.

Your task is to show that the identity holds.

Problem 4: Similar to the previous exercise we can compute the output of the softmax function $\pi_i = \frac{e^{x_i}}{\sum_{i=1}^N e^{x_i}}$ in a numerically stable way by shifting by an arbitrary constant a:

$$\frac{e^{x_i}}{\sum_{i=1}^N e^{x_i}} = \frac{e^{x_i - a}}{\sum_{i=1}^N e^{x_i - a}}$$

often chosen $a = \max_i x_i$. Show that the above identity holds.

F	rc	b	lei	m	3:	Ir	n	nac	hine	e le	arn	ing	yo	u c	fte	n c	om	ie a	ıcre	SS	pro	ble	ems	s w	hic	h c	ont	ain	th	e f	olle	wii	ng	qua	ant	ity							
		+						-												V																				-			
	_	+						+									y =	lo	g \	-	x_i																			\dashv			
		+																	i	1																				-			
F	or	· e	xa	m	ple	if y	ve :	wa	nt t	O C	alcı	ılat	e tl	he l	οg	-lik	elil	100	d c	of n	eur	alı	net	WO	rk v	vitl	h a	sof	tm	ax	out	DП	t w	e g	et.	this				-			
					-				ori						-																	-								-	-		
-									ws,																						_									-	-		
									enot		1 1														0 -		0	1	,					1						-	-		
		+	-		+			+	_	-																														-	-		
	O	CC	$^{ m pm}$	ıba	it t	his	ISS	sue	we	tyı	oica	lly	use	e ti	ie i	Ollo	OW)	ing	1d	ent	ity:																			-	-		
		+														-N						N																		\rightarrow			
		+											y		log	\sum	$\frac{1}{2}e^{i}$	r _i =	= a	+	log	\sum	e^x	i - a																\dashv			
		+														i=	1					i=1																		\rightarrow			
fe	or	aı	1 8	arb	oitr.	arv	a.	Т	his	me	ans	VO	011 C	an	shi	ft.	the	ce	nte	r o	f t.ŀ	ie e	XD	on ϵ	nti	al s	sum	1. A	\ tx	mic	cal	val	11e	is s	ett	ing)						
									a =																															-			
									ı ge										0																					\dashv			
		+					H.,	+	+																															\rightarrow			
									tha				- 1			ls.																											
		١.			N		ν:																																				
	6	-)		\geq	C	7-1																																	_			
		+			1=1					V		,																												\dashv			
		$^{+}$	d		Ч			- 0		>	P. 7																													\neg			
	Q	,	η.	(7	-	16	1	i	-1																																	
		T						1\/			(1 + x	,																															
			_	α	44		,	\$. 6) - (χι /	i																															
		4				_	i	FI		1																																	
				4							X	(-, -O	\																														
		Ŋ	-	9	=		100		\geq	(
									, = (/		2	ر. ₋	<u>_</u>																													
			Λ	١.	=	Q	+	1	ba	J. V.		P,																															
		(1-	-(
	P	ro	bl	lei					nila				_																			tl	nе	SO:	ftn	naz	x fi	unc	ctic	n	π_i	=	
	$^{\sim}$	e	i	x_i	ir	ı a	nı	ım.	eri	cal	ly s	sta	ble	W	ay	by	/ S	hif	tin	g l	ру	an	aı	bi	tra	ry	co	nst	an	t c	ι:												
	_	1=	-1																																								
																			e^{x}				$-\epsilon$	x_i	e^{x_i}																		
								-											V	e^x	i		\sim	V	ρx_i	-a														_			
		+						-										-i	=1				1=	=1														-					
	of	tε	\mathbf{n}	cl	os	en	a	⊨	ma	\mathbf{x}_i .	x_i .	Sl	ю	v t	ha	t t	he	ab	ΘV	e i	de:	nti	ty	$_{ m ho}$	lds					_										\rightarrow			
		+			- (> X	\			~ 0	(γ.	- A																								-	_		
	1	Λ	=		<u> `</u>	\	1		-		+		-	2	-	16	_//																							_			
	(}			\sum_{i}	7	P '	1	. 6	2				5	i= (ر م ا																								\rightarrow			
	_	+				ľ									' '																									-			
		+																																						\rightarrow			
																																								_			
		\dagger																																						\dashv	\dashv		
		+																																				\Box					

Problem 5: Load the notebook exercise_07_notebook.ipynb from Moodle. Fill in the missing code and run the notebook. Export (download) the evaluated notebook as PDF and add it to your submission.

We have implemented several helper functions for checking the correctness of your code in a small library nn_utils.py that can be downloaded from Moodle as well.

This week's programming assignment is closely connected to the contents of the in-class exercises. Make sure that you have a look at the in-class exercises before starting working on the homework task.

Note: We suggest that you use Anaconda for installing Python and Jupyter, as well as for managing packages. We recommend that you use Python 3.

For more information on Jupyter notebooks and how to convert them to other formats, consult the Jupyter documentation and nbconvert documentation.