Traditional Machine Learning

https://www.ai2lab.ca/

@lab_ai2

Revisiting traditional machine learning and how to combine it with deep learning models

Roberto Souza
Assistant Professor
Electrical and Software Engineering
Schulich School of Engineering

March 2022

Outline

- Machine Learning Recap
- Traditional Machine Learning
- Summary

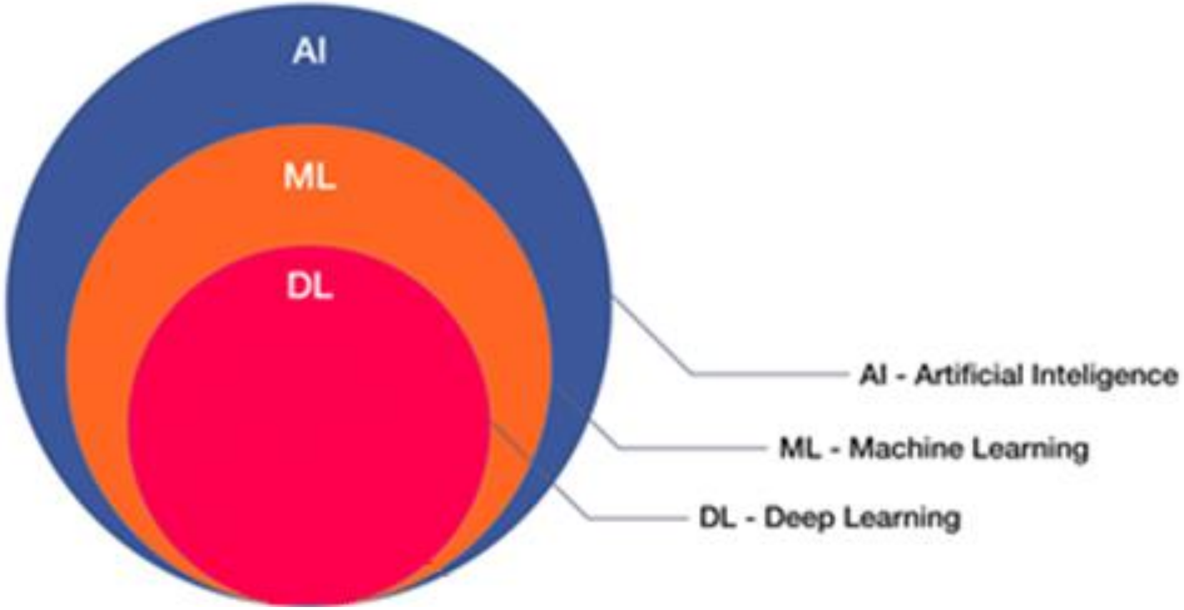
Learning goals

- An overview of how to deploy traditional machine learning models
- Learn strategies for combining deep learning features with handcrafted features
- Learn strategies for dealing with the curse of dimensionality

Machine Learning Recap

Artificial Intelligence (AI) Machine Learning (ML) and Deep Learning (DL)

- AI: the broad discipline of creating intelligent machines
- ML: refers to systems that can learn from experience
- DL: refers to systems that learn from experience on large data sets



ML techniques that do not fall under the DL category are often referred to as traditional ML.

Traditional ML

- Feature engineering
- “Simpler models” -> less parameters to be learned

$$\underbrace{X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ x_{31} & x_{32} & \dots & x_{3M} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{bmatrix}}_{\text{N samples with M features}} \qquad \underbrace{Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{bmatrix}}_{\text{Labels}}$$

Traditional ML libraries

scikits learn
machine learning in Python

scikit-image
image processing in python

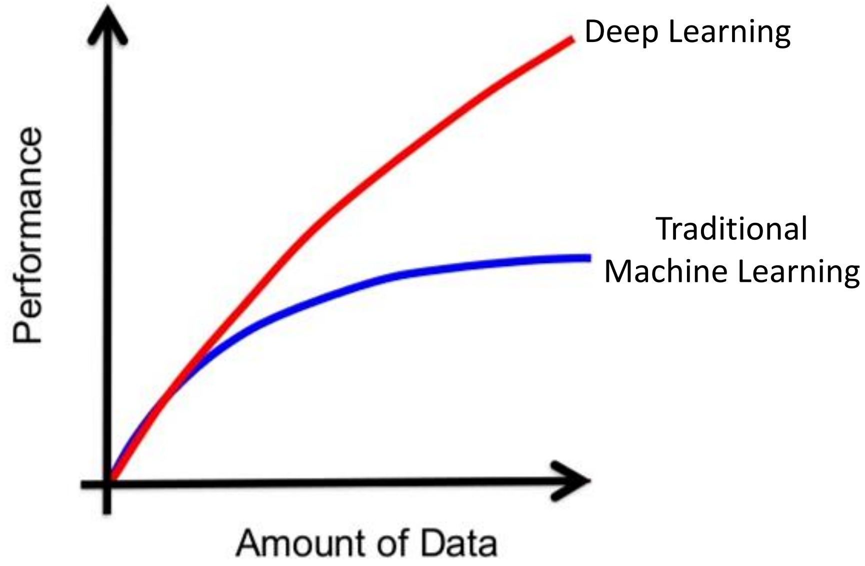
Imbalanced learn

pandas

Deep learning

- DL is a data-driven modeling approach, which “learns the features” -> input is raw data
  - But which features?

- Complex models with (tr, b, m)illions of parameters that need to be tuned



Frameworks accessible and easy to use

PYTORCH

- Supported by Facebook
- Preferred in academia for DL research

TensorFlow

- Supported by Google
- Stronger usage in industry
- With Keras, potentially easier to use

In practice, not much difference.

Traditional Machine Learning

Learn by Example

- Puzzle: As a credit company, it is important to know beforehand who is able to pay their loans and who is not. The goal of this puzzle is to build a statistical/machine learning model to figure out which clients can honor their debt.

ids	default	score_1	score_2	score_3	score_4	score_5	score_6	risk_rate	amount_b	borrowed	credit_lim	reason	income	sign	gender	facebook_	state	zip	channel	job_name	real_state	ok_since	n_bankru	n_default	n_account	n_issues
810e3277-	FALSE	smzX0nxh	tHpS8e9F	710	104.175	0.661509	123.0153	0.43	20024.31	60	62386	mLVIVxoG	59869.05	virg	f	TRUE	xsd3ZdsI3	i036nmJ7r	NCqL3QBx	mLVIVxoG	n+xK9CfX(	14	1	0	9	9
b4118fd5-	FALSE	DGCQep2	RO7MTL+j	330	97.8808	0.531115	110.9135	0.23	10046.51	36		mLVIVxoG	46016.31	sagi	f	FALSE	xsd3ZdsI3	oyrt7nHjo	NCqL3QBx	mLVIVxoG	n+xK9CfX(	75	0	0	3	
a75638f1-	FALSE	8k8UDR4Y	wkeCdGe	360	97.90893	0.611086	104.6208	0.3	21228.25	60		mLVIVxoG	48025.47	libr	m	TRUE	/L8vvVesE	BMIK35trN	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	5	
285ce334-	FALSE	4DLlLW62j	tQUTfUye	120	100.4346	0.139784	120.1347	0.15	23032.33	36		mLVIVxoG	46011.12		m	FALSE	GW2VZ3d	coa2oOrpj	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	5	
e643bf65-	FALSE	4DLlLW62j	7h8PTkrlT	330	103.7746	0.002856	104.3205	0.08	24026.29	36	32366	mLVIVxoG	90026.17	libr	m		sjJbkqJS7c	xTrDMEf/(	NCqL3QBx	mLVIVxoG	N5/CE7lSk	15	0	0	10	10
b84024d8-	FALSE	1Rk8w4Uc	rJZgTmAN	340	98.86923	0.424989	101.0563	0.22	8007.11	36		mLVIVxoG	43008.57		f	FALSE	1DpYl6dtz	ql9Tq+WR	NCqL3QBx	mLVIVxoG	n+xK9CfX(	13	0	0	16	16
8a66ed52-	FALSE	DGCQep2	RO7MTL+j	450	96.42699	0.396868	92.61916	0.42	10072.16	36	0	mLVIVxoG	39335.52	virg	m	FALSE	/EoxQEzy1	x0KGDyHi	NCqL3QBx0pscDnx3i		N5/CE7lSk	55	0	0	10	10
b8de2a5e	FALSE	smzX0nxh	bopP0NxV	450	97.79397	0.213243	82.37716	0.42	16016.97	60	102018	mLVIVxoG	98772		m	TRUE	kfFcVGcsJ	HNyKljS/K	NCqL3QBx	mLVIVxoG	N5/CE7lSk	17	0	0	13	13
63ada984-	FALSE	1Rk8w4Uc	dCm9hFKf	390	100.1189	0.123257	83.71872	0.36	11517.24	60	0	mLVIVxoG	62014.4		f	FALSE	BB/zpwTH	qf2kzehI0	NCqL3QBx	mLVIVxoG	N5/CE7lSk	49	0	0	6	6
08efdf0f-f	FALSE	DGCQep2	Fv28Bz0YF	280	95.39902	0.592517	101.6614	0.24	11539.08	36	82310	mLVIVxoG	85022.96		f	TRUE	BB/zpwTH	exaXk+hI3	NCqL3QBx	mLVIVxoG	N5/CE7lSk	43	0	0	19	19
acbb594c-	TRUE	DGCQep2	RO7MTL+j	250	95.75539	0.555542	98.94743	0.25	14036.89	36	9953	mLVIVxoG	40013.13	capr	m	TRUE	xsd3ZdsI3	vwoGcNrY	NCqL3QBx	mLVIVxoG	n+xK9CfX(	16	0	0	8	8
0854e48f-	FALSE	DGCQep2	osCzpM4h	300	102.8417	0.383545	90.58172	0.17	9020.12	36	25000	mLVIVxoG	37018.58	arie	m	FALSE	VafbDA6D	WYEA3Ulg	NCqL3QBx	mLVIVxoG	n+xK9CfX0bCn77lClT		0	0	15	15
8689dcdb-	FALSE	e4NYDor1	NLvAOzzn	120	97.39488	0.252639	89.94225	0.31	1418.09	36	0	mLVIVxoG	50015.28	libr	m	FALSE	BB/zpwTH	8REhu8TW	NCqL3QBx0pscDnx3i		n+xK9CfX0bCn77lClT		0	0	3	3
4322258a-	FALSE	DGCQep2	SaamrHM	480	98.57185	0.925637	98.58656	0.21	12018.15	36	81319	mLVIVxoG	72026.34	libr	m	FALSE	xsd3ZdsI3	vwoGcNrY	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	19	17
5e32067b-d9f9-2384-		e4NYDor1	cdpgyOyZ	330	103.4604	0.577916	74.00434	0.25	12409.15	60	22627	mLVIVxoG	49173.16		f	TRUE	QCVqG0w	A8Ltq8Y5D	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	7	7
739eec7d-ae35-dcd2		8k8UDR4Y	+CxEO4w7	520	94.67089	0.624899	110.3459	0.27	12319.32	60	48213	mLVIVxoG	35395.53	capr		TRUE	x6Gw/1I8t	Yw3BAzW	NCqL3QBx	mLVIVxoG	nSpvDsIsslUaX6GE6n		0	0	12	12
4c104001-	FALSE	4DLlLW62j	7h8PTkrlT	430	95.42024	0.130929	104.9058	0.12	8022.18	36	63601	mLVIVxoG	160033.7	arie	m	FALSE	x6Gw/1I8t	248/dJV3J	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	8	8
73e39a95-	FALSE	e4NYDor1	ky19q4V1	490	96.53734	0.664545	101.7721	0.33	2529	36		mLVIVxoG	40228.19	pisce	m	FALSE	JrdZzPZEa	7BAMPpe	NCqL3QBx	mLVIVxoG	n+xK9CfX(	18	0	0	19	19
95ba212b-	TRUE	DGCQep2	osCzpM4h	140	103.1767	0.503252	110.6148	0.17	19012.19	36		mLVIVxoG	43205.19		m	FALSE	xsd3ZdsI3	h4ONeP3	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	9	
958fa2a9-(	FALSE	DGCQep2	Fv28Bz0YF	400	102.3352	0.3009	92.74603	0.52	35019.04	36	67091	mLVIVxoG	95014.57	cance	m	FALSE	/+QaZYcpI	/3WrV3gr.	NCqL3QBx	mLVIVxoG	N5/CE7lSk	67	0	0	14	14
9fa432e4-	FALSE	DGCQep2	SaamrHM	310	98.40569	0.281539	101.8174	0.23	12021.03	36	54200	mLVIVxoG	62010.54		m	FALSE	kfFcVGcsJ	ceNlmpI8	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	11	11
b45e2ab5-	FALSE	4DLlLW62j	YLGMUI9h	350	98.32977	0.606256	97.22725	0.13	5021.16	36		mLVIVxoG	31893.23	pisce	f	TRUE	BB/zpwTH	EkqTGP51	NCqL3QBx	mLVIVxoG	nSpvDsIss	51	0	0	20	20
69f2c5da-(	FALSE	4DLlLW62j	tQUTfUye	420	98.9203	0.723942	104.6716	0.33	22084.74	36	55125	mLVIVxoG	70046.44	virg	m	FALSE	kfFcVGcsJ	JDf3u4+tjF	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	25	25
fc7ef4c5-b	FALSE	DGCQep2	7h+tk4z7C	310	99.39456	0.650101	120.862	0.34	21621.9	60		mLVIVxoG	72012.31	virg	f		orU7WJYG	5Op7K6Kc	NCqL3QBx	mLVIVxoG	n+xK9CfX0bCn77lClT		0	0	6	
e6fff980-5	FALSE	4DLlLW62j	YLGMUI9h	410	101.8274	0.728486	99.68445	0.15	20015.15	36		mLVIVxoG	154016.9		m	FALSE	x6Gw/1I8t	nP1q0F1u	NCqL3QBx	mLVIVxoG	n+xK9CfX(	42	0	0	7	
8e3422a9-	FALSE	1Rk8w4Uc	rJZgTmAN	170	97.53917	0.595424	91.30036	0.3	5021.59	36	0	mLVIVxoG	50010.83		m	FALSE	BB/zpwTH	8REhu8TW	NCqL3QBx	mLVIVxoG	n+xK9CfX0bCn77lClT		1	0	7	7
c81955ef-	FALSE	DGCQep2	RO7MTL+j	240	100.9793	0.588796	101.7241	0.2	18018.14	60		mLVIVxoG	121030.6	scor			ygE7OTds	4Hmfy977	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	9	9
ff860c5e-3	TRUE	8k8UDR4Y	mX2VRRG	240	101.3817	0.849653	90.29318	0.31	35017.06	36	41329	mLVIVxoG	121020.5	sagi	m	TRUE	xsd3ZdsI3	wXaZCM/I	NCqL3QBx	mLVIVxoG	n+xK9CfX(	39	0	0	10	10
7f93590f-4	TRUE	8k8UDR4Y	wkeCdGe	430	101.2879	0.694936	114.9513	0.61	35018.67	60		mLVIVxoG	80019.03	aqua	m	FALSE	xsd3ZdsI3	vjsYtmgFN	NCqL3QBx0pscDnx3i		N5/CE7lSkAfB04hVFl		0	0	10	
028a149e-	FALSE	DGCQep2	SaamrHM	290	101.4379	0.063472	90.18294	0.17	24017.18	60	56497	mLVIVxoG	98522.57	scor	m	FALSE	bNDXnbe	4UkFK78e	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	12	12
ac81ce99-	FALSE	1Rk8w4Uc	dCm9hFKf	290	97.38978	0.386192	90.79461	0.33	24608.49	60	0	mLVIVxoG	110039.8	gemi		TRUE	7bX4XFXn	27DtYF3F+	NCqL3QBx	mLVIVxoG	n+xK9CfX0bCn77lClT		0	0	7	7
7b1e275e-	FALSE	DGCQep2	RO7MTL+j	310	100.9584	0.600512	98.29693	0.48	21861.88	60	0	mLVIVxoG	50024.71		m	FALSE	bwNrC22V	vYKZVZfA	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	9	9
e76c6fa8-	FALSE	8k8UDR4Y	d/7Hedyz	330	105.2137	0.265397	104.5784	0.38	7015.87	36	34039	mLVIVxoG	75015.06	capr	f	TRUE	1vMmtGU	HqN+6NO	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	5	5
77a10088-	FALSE	4DLlLW62j	pAzpxkhjF	340	93.87215	0.777031	103.0404	0.3	11017.37	36		mLVIVxoG	100007.7	libr	m	FALSE	1DpYl6dtz	8qgxecfb\	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		0	0	8	
26a68732-	FALSE	DGCQep2	SaamrHM	470	103.5186	0.082595	102.4772	0.25	18018.63	36	34819	mLVIVxoG	64032.34	gemi	m	TRUE	xsd3ZdsI3	YA0xaBEZ	NCqL3QBx	mLVIVxoG	n+xK9CfX(	16	0	0	10	9
fd228aa8-	FALSE	4DLlLW62j	pAzpxkhjF	360	102.9325	0.980142	109.1629	0.17	14016.94	36	34558	mLVIVxoG	65035.38	cance	m	FALSE	fR2lwAMj	rvdltYNoj	NCqL3QBx	mLVIVxoG	n+xK9CfX(	19	0	0	20	20
083ca221-	FALSE	DGCQep2	osCzpM4h	280	102.8746	0.561958	114.914	0.21	7163.14	36	0	mLVIVxoG	31039.94	scor	m	FALSE	82aTqSzrT	ap0+SDtW	NCqL3QBx	mLVIVxoG	N5/CE7lSkAfB04hVFl		1	0	4	4

Step-by-step Analysis

1. Exploratory Data Analysis (EDA)
2. Experimental Setup
   - Data imputation
   - Data split
   - Metrics

- Model Selection
- Analysis of features importances’

Combining Deep Learning and Machine Learning

- Some times it is worth to combine deep learning features with traditional machine learning models
- Combining deep learning features with handcrafted features
  - Can increase accuracy of the models
  - Make models more explainable
- The curse of dimensionality
  - Global Average Pooling
  - Dimensionality reduction
    - Principal Component Analysis
    - Independent Component Analysis

Summary

Thank you!
Questions?