



Universidade Federal do Ceará

Campus de Itapajé

Redes Neurais Artificiais

TRABALHO COMPUTACIONAL
Mini-projeto I:

Prof. Dr. Hitalo Nascimento
hitalo.nascimento@ufc.br



Projeto 1

Esse projeto consiste na implementação de uma RNA do tipo MLP, visando a análise de um problema de classificação de três **cultivares** de vinhos diferentes na Itália. Nesse sentido, um data set contendo os resultados da análise físico-química de 178 amostras de vinho tinto de tais cultivares é fornecido para realização do projeto.



Projeto 1

Ao todo, 13 parâmetros físico-químicos estão disponíveis no data set:

<https://archive.ics.uci.edu/dataset/109/wine>

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline



Descrição

- Exemplo do data set. A primeira coluna, refere-se à classe.

1	1	14.23	1.71	2.43	15.6	127	2.8	3.06	.28	2.29	5.64	1.04	3.92	1065
2	1	13.2	1.78	2.14	11.2	100	2.65	2.76	.26	1.28	4.38	1.05	3.4	1050
3	1	13.16	2.36	2.67	18.6	101	2.8	3.24	.3	2.81	5.68	1.03	3.17	1185
4	1	14.37	1.95	2.5	16.8	113	3.85	3.49	.24	2.18	7.8	.86	3.45	1480
5	1	13.24	2.59	2.87	21	118	2.8	2.69	.39	1.82	4.32	1.04	2.93	735
6	1	14.2	1.76	2.45	15.2	112	3.27	3.39	.34	1.97	6.75	1.05	2.85	1450
7	1	14.39	1.87	2.45	14.6	96	2.5	2.52	.3	1.98	5.25	1.02	3.58	1290
8	1	14.06	2.15	2.61	17.6	121	2.6	2.51	.31	1.25	5.05	1.06	3.58	1295



Projeto 2

Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo da classificação é prever se o cliente irá realizar (sim/não) um depósito bancário (variável y). Nesse sentido, os dados são oriundos de campanhas de marketing de uma instituição bancária portuguesa. As campanhas de marketing foram baseadas em ligações telefônicas.



Projeto 2

Ao todo, 17 parâmetros estão disponíveis no data set:
<https://archive.ics.uci.edu/data-set/222/bank+marketing>

Variable Name	Role	Type	Demographic	Description
age	Feature	Integer	Age	
job	Feature	Categorical	Occupation	type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','m employed','services','student','technician','unemployed','unknown')
marital	Feature	Categorical	Marital Status	marital status (categorical: 'divorced','married','single','unknown'; note: 'div widowed')
education	Feature	Categorical	Education Level	(categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','u
default	Feature	Binary		has credit in default?
balance	Feature	Integer		average yearly balance
housing	Feature	Binary		has housing loan?
loan	Feature	Binary		has personal loan?
contact	Feature	Categorical		contact communication type (categorical: 'cellular','telephone')
day_of_week	Feature	Date		last contact day of the week
month	Feature	Date		last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
duration	Feature	Integer		last contact duration, in seconds (numeric). Important note: this attribute h target (e.g., if duration=0 then y='no'). Yet, the duration is not known befor after the end of the call y is obviously known. Thus, this input should only b purposes and should be discarded if the intention is to have a realistic pred
campaign	Feature	Integer		number of contacts performed during this campaign and for this client (nur
pdays	Feature	Integer		number of days that passed by after the client was last contacted from a pr -1 means client was not previously contacted)
previous	Feature	Integer		number of contacts performed before this campaign and for this client
poutcome	Feature	Categorical		outcome of the previous marketing campaign (categorical: 'failure','nonexis
y	Target	Binary		has the client subscribed a term deposit?



Projeto 2

Exemplo do data set. A última coluna, refere-se a classe.

"age"	"job"	"marital"	"education"	"default"	"balance"	"housing"	"loan"	"contact"	"day"	"month"	"duration"	"campaign"	"pdays"	"previous"	"poutcome"	"y"
58	"management"	"married"	"tertiary"	"no"	2143	"yes"	"no"	"unknown"	5	"may"	261	1	-1	0	"unknown"	"no"
44	"technician"	"single"	"secondary"	"no"	29	"yes"	"no"	"unknown"	5	"may"	151	1	-1	0	"unknown"	"no"
33	"entrepreneur"	"married"	"secondary"	"no"	2	"yes"	"yes"	"unknown"	5	"may"	76	1	-1	0	"unknown"	"no"
47	"blue-collar"	"married"	"unknown"	"no"	1506	"yes"	"no"	"unknown"	5	"may"	92	1	-1	0	"unknown"	"no"
33	"unknown"	"single"	"unknown"	"no"	1	"no"	"no"	"unknown"	5	"may"	198	1	-1	0	"unknown"	"no"
35	"management"	"married"	"tertiary"	"no"	231	"yes"	"no"	"unknown"	5	"may"	139	1	-1	0	"unknown"	"no"
28	"management"	"single"	"tertiary"	"no"	447	"yes"	"yes"	"unknown"	5	"may"	217	1	-1	0	"unknown"	"no"
42	"entrepreneur"	"divorced"	"tertiary"	"yes"	2	"yes"	"no"	"unknown"	5	"may"	380	1	-1	0	"unknown"	"no"
58	"retired"	"married"	"primary"	"no"	121	"yes"	"no"	"unknown"	5	"may"	50	1	-1	0	"unknown"	"no"
43	"technician"	"single"	"secondary"	"no"	593	"yes"	"no"	"unknown"	5	"may"	55	1	-1	0	"unknown"	"no"
41	"admin."	"divorced"	"secondary"	"no"	270	"yes"	"no"	"unknown"	5	"may"	222	1	-1	0	"unknown"	"no"



Projeto 3

Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo é classificar pessoas como tendo crédito bom ou ruim. Dois data sets são fornecidos. O **data set original**, no formato fornecido pelo Prof. Hofmann, que contém atributos categóricos/simbólicos e está no arquivo "**german.data**". Para algoritmos que necessitam de atributos numéricos, pode-se usar o arquivo "**german.data-numeric**". Este arquivo foi editado para torná-lo adequado para algoritmos que não conseguem lidar com variáveis categóricas. Vários atributos ordenados categóricos (como o atributo 17) foram codificados como inteiros.



Projeto 3

Ao todo, 20 parâmetros estão disponíveis no data set :

<https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data>

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Attribute1	Feature	Categorical		Status of existing checking account		no
Attribute2	Feature	Integer		Duration	months	no
Attribute3	Feature	Categorical		Credit history		no
Attribute4	Feature	Categorical		Purpose		no
Attribute5	Feature	Integer		Credit amount		no
Attribute6	Feature	Categorical		Savings account/bonds		no
Attribute7	Feature	Categorical	Other	Present employment since		no
Attribute8	Feature	Integer		Installment rate in percentage of disposable income		no
Attribute9	Feature	Categorical	Marital Status	Personal status and sex		no
Attribute10	Feature	Categorical		Other debtors / guarantors		no
Attribute11	Feature	Integer		Present residence since		no
Attribute12	Feature	Categorical		Property		no
Attribute13	Feature	Integer	Age	Age	years	no
Attribute14	Feature	Categorical		Other installment plans		no
Attribute15	Feature	Categorical	Other	Housing		no
Attribute16	Feature	Integer		Number of existing credits at this bank		no
Attribute17	Feature	Categorical	Occupation	Job		no
Attribute18	Feature	Integer		Number of people being liable to provide maintenance for		no
Attribute19	Feature	Binary		Telephone		no
Attribute20	Feature	Binary	Other	foreign worker		no
class	Target	Binary		1 = Good, 2 = Bad		no



Projeto 3

Exemplo do dataset. A última coluna, refere-se a classe (1 – BOM / 2 – RUIM).

1	1	6	4	12	5	5	3	4	1	67	3	2	1	2	1	0	0	1	0	0	1	0	0	1	1
2	2	48	2	60	1	3	2	2	1	22	3	1	1	1	1	0	0	1	0	0	1	0	0	1	2
3	4	12	4	21	1	4	3	3	1	49	3	1	2	1	1	0	0	1	0	0	1	0	1	0	1
4	1	42	2	79	1	4	3	4	2	45	3	1	2	1	1	0	0	0	0	0	0	0	0	1	1
5	1	24	3	49	1	3	3	4	4	53	3	2	2	1	1	1	0	1	0	0	0	0	0	1	2
6	4	36	2	91	5	3	3	4	4	35	3	1	2	2	1	0	0	1	0	0	0	0	1	0	1
7	4	24	2	28	3	5	3	4	2	53	3	1	1	1	1	0	0	1	0	0	1	0	0	1	1
8	2	36	2	69	1	3	3	2	3	35	3	1	1	2	1	0	1	1	0	1	0	0	0	0	1
9	4	12	2	31	4	4	1	4	1	61	3	1	1	1	1	0	0	1	0	0	1	0	1	0	1
10	2	30	4	52	1	1	4	2	3	28	3	2	1	1	1	1	0	1	0	0	1	0	0	0	2



Projeto 4

Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo da classificar estudantes em umas das seguintes categorias (abandono, matrícula e graduado / formando) ao final da duração normal de um curso de graduação.

- O data set foi criado a partir de uma instituição de ensino superior, para alunos matriculados em diferentes cursos de graduação, como agronomia, design, educação, enfermagem, jornalismo, gestão, serviço social e tecnologias.



Projeto 4

O data set inclui informações conhecidas no momento da matrícula dos alunos (percurso acadêmico, dados demográficos e fatores socioeconômicos) e o desempenho acadêmico dos alunos no final do primeiro e segundo semestres.

Ao todo, 36 parâmetros estão disponíveis no banco de dados <https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>



Projeto 4

Exemplo do data set. A última coluna, refere-se a classe.

<u>Curricular units 2nd sem (without evaluations)</u>	<u>Unemployment rate</u>	<u>Inflation rate</u>	<u>GDP</u>	<u>Target</u>
0	10.8	1.4	1.74	Dropout
0	13.9	-0.3	0.79	Graduate
0	10.8	1.4	1.74	Dropout
0	9.4	-0.8	-3.12	Graduate
0	13.9	-0.3	0.79	Graduate
5	16.2	0.3	-0.92	Graduate
0	15.5	2.8	-4.06	Graduate
0	15.5	2.8	-4.06	Dropout
0	16.2	0.3	-0.92	Graduate
0	8.9	1.4	3.51	Dropout
0	13.9	-0.3	0.79	Graduate



Projeto 5

- Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo da classificação é prever se uma casal irá se divorciar (SIM ou NÃO). Trata-se um conjunto de dados real, construído a partir de um formulário com 170 respostas.
- Ao todo, 54 parâmetros estão disponíveis no banco de dados <https://archive.ics.uci.edu/dataset/497/divorce+predictors+data+set>

[illegible]



Projeto 6

- Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo é classificar cerâmicas com base em sua composição química;
- Ao todo, 19 parâmetros estão disponíveis no seguinte conjunto de dados:<https://archive.ics.uci.edu/dataset/583/chemical+composition+of+ceramic+samples>



Projeto 6

Ceramic.Name: name of ceramic types from Longquan and Jindgezhen

Part: a binary categorical variable ('Body' or 'Glaze')

Na2O: percentage of Na2O (wt%)

MgO: percentage of MgO (wt%)

Al2O3: percentage of Al2O3 (wt%)

SiO2: percentage of SiO2 (wt%)

K2O: percentage of K2O (wt%)

CaO: percentage of CaO (wt%)

TiO2: percentage of TiO2 (wt%)

Fe2O3: percentage of Fe2O3 (wt%)

MnO: percentage of MnO (ppm)

CuO: percentage of CuO (ppm)

ZnO: percentage of ZnO (ppm)

PbO2: percentage of PbO2 (ppm)

Rb2O: percentage of Rb2O (ppm)

SrO: percentage of SrO (ppm)

Y2O3: percentage of Y2O3 (ppm)

ZrO2: percentage of ZrO2 (ppm)

P2O5: percentage of P2O5 (ppm)



Projeto 6

Exemplo do data set. A primeira coluna, refere-se a classe.

1	<u>Ceramic Name</u>	<u>Part</u>	<u>Na2O</u>	<u>MgO</u>	<u>Al2O3</u>	<u>SiO2</u>	<u>K2O</u>	<u>CaO</u>	<u>TiO2</u>	<u>Fe2O3</u>
2	<u>FLO-1-b</u>	<u>Body</u>	0.62	0.38	19.61	71.99	4.84	0.31	0.07	1.18
3	<u>FLO-2-b</u>	<u>Body</u>	0.57	0.47	21.19	70.09	4.98	0.49	0.09	1.12
4	<u>FLO-3-b</u>	<u>Body</u>	0.49	0.19	18.60	74.70	3.47	0.43	0.06	1.07
5	<u>FLO-4-b</u>	<u>Body</u>	0.89	0.30	18.01	74.19	4.01	0.27	0.09	1.23
6	<u>FLO-5-b</u>	<u>Body</u>	0.03	0.36	18.41	73.99	4.33	0.65	0.05	1.19
7	<u>FLO-6-b</u>	<u>Body</u>	0.62	0.18	18.82	73.79	4.28	0.30	0.04	0.96
8	<u>FLO-7-b</u>	<u>Body</u>	0.45	0.33	17.65	74.99	3.53	0.70	0.07	1.28
9	<u>FLO-8-b</u>	<u>Body</u>	0.50	0.45	21.40	71.40	3.47	0.35	0.05	1.00



Projeto 7

- Esse projeto consiste na implementação de uma RNA do tipo MLP, cuja o objetivo é prever a idade de um abalone a partir de medições físicas;
- Ao todo, 8 parâmetros estão disponíveis no seguinte conjunto de dados: <https://archive.ics.uci.edu/dataset/1/abalone>



Projeto 7

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Sex	Feature	Categorical		M, F, and I (infant)		no
Length	Feature	Continuous		Longest shell measurement	mm	no
Diameter	Feature	Continuous		perpendicular to length	mm	no
Height	Feature	Continuous		with meat in shell	mm	no
Whole_weight	Feature	Continuous		whole abalone	grams	no
Shucked_weight	Feature	Continuous		weight of meat	grams	no
Viscera_weight	Feature	Continuous		gut weight (after bleeding)	grams	no
Shell_weight	Feature	Continuous		after being dried	grams	no
Rings	Target	Integer		+1.5 gives the age in years		no



Projeto 7

Exemplo do data set. A última coluna, refere-se a classe.

```
1M,0.455,0.365,0.095,0.514,0.2245,0.101,0.15,15
2M,0.35,0.265,0.09,0.2255,0.0995,0.0485,0.07,7
3F,0.53,0.42,0.135,0.677,0.2565,0.1415,0.21,9
4M,0.44,0.365,0.125,0.516,0.2155,0.114,0.155,10
5I,0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7
6I,0.425,0.3,0.095,0.3515,0.141,0.0775,0.12,8
7F,0.53,0.415,0.15,0.7775,0.237,0.1415,0.33,20
```



Descrição

Em todos os projetos, a Rede Neural deve ser treinada até que consiga atingir uma acurácia de pelo menos 80% em relação aos testes.

Parte I: Defina e implemente os componentes necessários para o treinamento da rede neural:

- **Entradas;**
- **Pesos:** Utilizar a inicialização de Xavier;
- **Duas camadas escondidas:** em que o número de unidades de processamento é igual ao tamanho da entrada;
- **N saídas:** a depender do problema; taxa de aprendizado; função de custo; o número de iterações e épocas);



Descrição

Parte II: Implemente a propagação para frente e calcule o custo;

Parte III: Calcule o gradiente da função de custo;

Parte IV: Implemente a retropropagação. (Use a regra de atualização para o gradiente descendente);

Parte V: Teste diferentes números de camadas intermediárias, e verifique se há uma melhora na qualidade dos resultados no teste. Qual o número de camadas que ideal?



Descrição

- O trabalho pode ser feito em dupla;
- Deve ser apresentado em sala de aula até o dia 09/11;
- Deve ser implementado em qualquer linguagem de programação;
- É necessário evidenciar a execução do código.