





/ Categorical Encoding



Categorical Features

/ Examples of categorical features (aka nominal features):

- Name of the person
- Country of the person
- Etc.



Common categorical encodings

Ordinal encoding

Categorical Feature		Numeric
Louise	=>	1
Gabriel	=>	2
Emma	=>	3
Adam	=>	4
Alice	=>	5
Raphael	=>	6
Chloe	=>	7
Louis	=>	8
Jeanne	=>	9
Arthur	=>	10

Useful for tree models
(Random Forest, GBMs)

Binary encoding

		Binary Encoded			
Categorical Feature	=	x1	x2	x4	x8
Louise	=>	1	0	0	0
Gabriel	=>	2	0	1	0
Emma	=>	3	1	1	0
Adam	=>	4	0	0	1
Alice	=>	5	1	0	1
Raphael	=>	6	0	1	1
Chloe	=>	7	1	1	1
Louis	=>	8	0	0	0
Jeanne	=>	9	1	0	0
Arthur	=>	10	0	1	0

One Hot Encoding (Best option)

Categorical Feature		f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
Louise	=>	1	0	0	0	0	0	0	0	0	0
Gabriel	=>	0	1	0	0	0	0	0	0	0	0
Emma	=>	0	0	1	0	0	0	0	0	0	0
Adam	=>	0	0	0	1	0	0	0	0	0	0
Alice	=>	0	0	0	0	1	0	0	0	0	0
Raphael	=>	0	0	0	0	0	1	0	0	0	0
Chloe	=>	0	0	0	0	0	0	1	0	0	0
Louis	=>	0	0	0	0	0	0	0	1	0	0
Jeanne	=>	0	0	0	0	0	0	0	0	1	0
Arthur	=>	0	0	0	0	0	0	0	0	0	1

Useful for multiplicative models
(Linear models, Neural Nets, SVMs)



/ Q&A

What are your doubts?

