



100 Days of ML

- Day 9 - MLDLC
- Day 10 - Job Roles
- Day 11 - Tensors
- Day 12 - Setting u...
- Day 13 - End to En...
- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...

+ Add section + Add page

OneHotEncoding

Tuesday, April 13, 2021 12:24 PM

Color	Target
Yellow	0
Yellow	1
Blue	1
Yellow	1
Red	1
Yellow	0
Red	1
Red	0
Yellow	1
Blue	0

$[Y, B, R]$
0 1 2

Color_Y	Color_B	Color_R	Target
1	0	0	0
1	0	0	1
0	1	0	1
1	0	0	1
0	0	1	1
1	0	0	0
0	0	1	1
0	0	1	0
1	0	0	1
0	1	0	0

One Hot Encoding

$[1, 0, 0] \rightarrow \text{Yellow}$

$[0, 1, 0] \rightarrow \text{Blue}$

$[0, 0, 1] \rightarrow \text{Red}$





Dummy Variable Trap

Tuesday, April 13, 2021 12:26 PM

Color	Target
Yellow	0
Yellow	1
Blue	1
Yellow	1
Red	1
Yellow	0
Red	1
Red	0
Yellow	1
Blue	0

n cat
($n-1$)
cols

Color_Y	Color_B	Color_R	Target
1	0	0	0
1	0	0	1
0	1	0	1
1	0	0	1
0	0	1	1
1	0	0	0
0	0	1	1
0	0	1	0
1	0	0	1
0	1	0	0

n

n cols

$(n-1)$ cols

One Hot Encoding

multicollinearity

$Y -$
 $B -$
 $R -$

$\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\sum = 1$

linear model

Linear Reg

Logistic



← → OneNote for Windows 10

Home Insert Draw View Help

↶ ↷ AI + ⇄

📁 100 Days of ML ▾

- Day 9 - MLDLC
- Day 10 - Job Roles
- Day 11 - Tensors
- Day 12 - Setting u...
- Day 13 - End to En...
- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...

OneHotEncoding

Dummy Variable Trap

OHE using most frequent variables

Example

+ Add section + Add page

OHE using most frequent variables

Tuesday, April 13, 2021 12:31 PM

brand → nominal → 40 brands
cat → OHE

| | | | |

□

frequent cat (10)
↳ others

