

Hash2Vec

Encoding Features

```
i make ten thousand dollars per week just surfing the  
are you free for a meeting early next week?
```

There are 19 unique words in total, each is assigned
a unique index.

```
i: 0  
make: 1  
ten: 2  
thousand: 3  
dollars: 4  
per: 5  
week: 6  
just: 7  
surfing: 8  
the: 9  
web: 10  
are: 11  
you: 12  
free: 13  
for: 14  
a: 15  
meeting: 16  
early: 17  
next: 18
```

Hash2Vec

There are many hash functions available, but they all do the same thing:

Map data of arbitrary size to data of a fixed size.

"John Doe" -> hash function -> 34

"Jane Doe" -> hash function -> 48

word2vec

- Continuous bag of words model

Why not word2vec?

Curse of dimensionality

Memory consumption

More complex

Hash2vec

Feature Engineering
Explain with example
Cons

Thank You!