

Modeling

- Vector space
- Features are dimensions
- Feature extraction
- Feature engineering (synthetic features)
 - Brainstorm
 - Pick some
 - Make them
 - Evaluate
 - Repeat
- [titanic titles](#)
- [reading-time](#)
- one-hot encoding
- [python/pandas examples](#)

Example: Text Features

- Bag of words
 - corpus (documents)
 - vocabulary (set of unique words)
 - words
- Notes
 - order doesn't matter
 - stop words
 - stemming (racinisation, désuffixation)
 - Lemmatisation (transformer en lemme)

Example: Image Features

- Corners, edges (rotation invariant, but scaling can hide)
- More complex: scale space or RNN
- Point matching is easy
- Problems
 - Illumination
 - Scale
 - Rotation
 - Skew (perspective)
 - Data size (matrices not sparse)

Linear Regression

[slides](#)