

# 自然言語処理入門

岸山 健 (31-187002)

Nov. 19, 2018

## 課題

データを用いて「空気」の語義がどちらであることを推定せよ。

- 水を含んだ湿った空気が流れてくる (空気 1)
- 時代を覆う湿った空気が読み取れる (空気 2)

第五回の spam/ham の枠組み, ナイーブ・ベイズを用いて上のデータがどちらの空気クラス C (空気 1(地球を含んでいる無色透明の気体), 空気 2(その場の気分, 雰囲気)) に属するかを考える. 今回の場合は 2 クラスなので, 「空気 1 である確率」を「空気 2 である確率」で割りオッズ比を求める. その場合, 与えられた素性において空気 1 である確率のほうが高ければ, オッズ比は 1 より大きくなる. なお, 与えられた単語の出現が独立であると仮定する.

まずは最初の例のオッズは  $p > 1$  であるため, 語義は空気 1(地球を含んでいる無色透明の気体) となる.

```
(0.48*0.08*0.122*0.026) / (0.52*0.007*0.002*0.048)
# [1] 348.5714
```

次の例のオッズは  $p < 1$  であるため, 語義は空気 2(その場の気分, 雰囲気) となる.

```
(0.48*0.014*0.012*0.122*0.005) / (0.52*0.07*0.073*0.002*0.196)
# [1] 0.04722479
```

いずれも直感とあう結果である.