Data For Machine Learning

E.g. Classify btw confusable words.

"For breakfast I ate ___ eggs.

{ to, two, too }

Algorithms
- Perceptron   - Memory-based
- Winnow   - Naïve Bayes

"It's not who has the best algorithm that wins.
It's who has the most data."

Large data rationale

feature $x \in \mathbb{R}^{n+1}$ 이 y 를 정확히 예측할

충분한 정보를 가진다고 가정

- Use a learning algorithm with many parameters (low bias)   복잡
- Use a very large training set (unlikely to overfit) (low var)

↳ 자 information이 빈약하면 training set 크기

키도 improvement 없음.