



Why Language is Hard: Structure and Predictions

Advanced Machine Learning for NLP

Jordan Boyd-Graber

INTRODUCTION

Today

- What's special about machine learning for NLP
- Layout of the course
- Administrivia
- Perceptron
- Structured Perceptron

Today

- What's special about machine learning for NLP
- Layout of the course
- Administrivia
- Perceptron
- Structured Perceptron
 - ① Good ML analysis, standard NLP problem
 - ② Often ignored in both classes (except when I teach it)
 - ③ Uses structure and representation

Most supervised algorithms are ...

Logistic Regression

Most supervised algorithms are ...

Logistic Regression

$$p(y|x) = \sigma(\sum_i \beta_i x_i)$$

SVM

Most supervised algorithms are ...

Logistic Regression

$$p(y|x) = \sigma(\sum_i \beta_i x_i)$$

SVM

$$\text{sign}(\vec{w} \cdot x + b)$$

- What statistical property do these (and many others share)?

Most supervised algorithms are ...

Logistic Regression

$$p(y|x) = \sigma(\sum_i \beta_i x_i)$$

SVM

$$\text{sign}(\vec{w} \cdot x + b)$$

- What statistical property do these (and many others share)?
- Hint: $p(y_i, y_j | x_i, x_j) = p(y_i | x_i)p(y_j | x_j)$

Most supervised algorithms are ...

Logistic Regression

$$p(y|x) = \sigma(\sum_i \beta_i x_i)$$

SVM

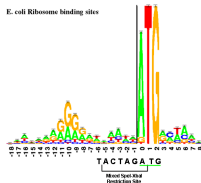
$$\text{sign}(\vec{w} \cdot x + b)$$

- What statistical property do these (and many others share)?
- Hint: $p(y_i, y_j | x_i, x_j) = p(y_i | x_i)p(y_j | x_j)$
- Independent!

Is this how the world works?



E. coli Ribosome binding sites



Also particularly relevant for 2016: correlated voting patterns

Why is this ML class different from all other ML classes?

- NLP has very specific applications
- NLP has very specific ML problems
- Much of the skills you need to do ML well are domain-specific
- Culture in ML for NLP research is slightly different than vanilla ML
 - Cleverness is not enough
 - Good baselines are important
 - Simple is usually better