

Introducción a Natural Language Processing

Modelos Discriminativos

Ronald Cárdenas Acosta

Setiembre, 2016

Outline

1 Definición

2 Features en NLP

Modelos Discriminativos

- Recordatorio: los modelos generativos estiman las distribuciones de *Class Prior*, $P(Y)$, y *Class Conditionals*, $P(X|Y)$.
- Cualquier modelo que no estime estas distribuciones, es llamado *Discriminativo*.
- Modelar el proceso de generación de la data es útil, pero no siempre necesario.
- Un clasificador es cualquier función que mapee objetos $x \in \mathcal{X}$ hacia clases $y \in \mathcal{Y}$.

Features en NLP

- Una característica o *feature* es un número real que describe una propiedad del objeto x .
- En NLP, éstas por lo general son binarias (0/1) y son el resultado de evaluar proposiciones como:

$$g_1(x) = \begin{cases} 1 & \text{si la oración } x \text{ contiene la palabra } \textit{Farfán} \\ 0 & \text{si no} \end{cases}$$

$$g_2(x) = \begin{cases} 1 & \text{si } x \text{ tiene las palabras } \textit{excelente}, \textit{exitoso} \\ 0 & \text{si no} \end{cases}$$

- La representación vectorial de características de x está dada por:

$$g(x^i) = [g_1(x^i), g_2(x^i), \dots, g_M(x^i)]$$

Features en NLP

- Para multi-clasificación, es útil asociar *features* con el objeto de entrada y su respectiva clase
- Se define la función de mapeo conjunto (*joint feature mapping*):
 $f : X \times Y \rightarrow R^M$
- En este caso, la representación vectorial de características de x está dada por:

$$f(x^i, y^i) = [f_1(x^i, y^i), f_2(x^i, y^i), \dots, f_M(x^i, y^i)]$$

$$f_1(x, y) = \begin{cases} 1 & \text{si } x \text{ contiene Farfán, y el t\u00f3pico } y \text{ es } \textit{Deportes} \\ 0 & \text{si no} \end{cases}$$

$$f_2(x, y) = \begin{cases} 1 & \text{si } x \text{ contiene Farfán, y el t\u00f3pico } y \text{ es } \textit{Pol\u00edtica} \\ 0 & \text{si no} \end{cases}$$

Inferencia o Predicción

- Para un modelo lineal, se tiene:

$$\hat{y} = \operatorname{argmax}_{y \in Y} w \cdot f(x, y)$$

- Cada clase esta asociada un vector de parámetros w_y , de manera que:

$$w \cdot f(x, y) = w_y \cdot g(x)$$