

# Adversarial Training with Contrastive Learning in NLP

Máster en Ciencia de Datos

---

Francisco Javier Sáez Maldonado

10 de marzo de 2022

Procesamiento de Lenguaje Natural

*Escuela Politécnica Superior  
Universidad Autónoma de Madrid*

## 1. Teoría de la información

Información mutua

Cotas inferiores

## 2. Aprendizaje contrastivo

Estimación del ruido contrastiva

Contrastive predictive coding

Pérdida usando tripletas

## 3. Nuevos marcos de trabajo

SimCLR

Bootstrap your own latent

## 4. Experimentación

Objetivos

Experimentos con SimCLR

Experimentos con BYOL

**Dato**

(0.1, 0, 2, 1, 0, 0.5, 2.4, 5)

**Etiqueta**

Perro

**Dato**

(0.1, 0, 2, 1, 0, 0.5, 2.4, 5)

**Etiqueta**

Perro

Sea  $x \in \mathbb{R}^d$  un vector de entrada a un modelo de aprendizaje automático. Una *representación*  $\tilde{x} \in \mathbb{R}^n$  es otro vector de menor dimensión que comparte información o características con  $x$ .

**Objetivo:** extraer **representaciones** que sean buenas en general para **tareas posteriores**.

# Introducción

---

## **Aprendizaje contrastivo**

---

### Definición (Pérdida contrastiva)

Sea  $X = \{x^*, x_1, \dots, x_{N-1}\}$  un conjunto de  $N$  ejemplos donde  $x^*$  ha sido extraído de la distribución conjunta  $P(x, z)$  y le resto han sido extraídos del producto de las distribuciones marginales  $P(x), P(z)$ . Se define entonces la función de pérdida contrastiva como

$$\ell(\theta) = -E_X \left[ \log \frac{h_\theta(x^*, z)}{\sum_{x \in X} h_\theta(x, z)} \right].$$

## Tercer experimento y resultados finales





- El uso de la teoría de la información proporciona un buen punto de partida para el aprendizaje de representaciones.
- El aprendizaje contrastivo ha probado ser la mejor forma de obtener representaciones que son útiles en tareas posteriores.
- Ambos marcos de trabajo probados obtienen buenos resultados en la adaptación a conjuntos de datos más pequeños.

**Gracias por su atención**