



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO

## Aprendizaje de transferencia

Macroentrenamiento en Inteligencia Artificial (MeIA)

Dr. Magdiel Jiménez Guarneros

20 de Junio de 2023





- ▶ **Introducción**
- ▶ Desplazamiento del conjunto de datos
- ▶ Aprendizaje de transferencia
- ▶ Adaptación de dominio

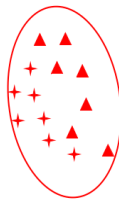


Dos suposiciones son críticas para la eficacia de métodos de modelos de aprendizaje supervisado

- **Cantidad de datos:** Disponibilidad de suficientes datos de entrenamiento, preferiblemente etiquetados.
- **Calidad de los datos:** Los datos de entrenamiento y prueba obedecen a la condición de una distribución independiente e idéntica.



Entrenamiento



Prueba



En la aplicaciones reales es difícil satisfacer las dos suposiciones debido a diferentes factores

- **Variabilidad entre sujetos:** Existe una alta variabilidad de las señales de EEG entre sujetos debido diferencias fisiológicas y funcionales.
- **Variabilidad entre sesiones:** Las señales de EEG del mismo sujeto pueden variar incluso bajo para la misma actividad cerebral en diferentes momentos, lugares y escenas.
- **Interferencia de ruido:** las señales de EEG se distorsionan fácilmente no solo por los ruidos instrumentales y ambientales, sino también por las interferencias físicas y mentales.

El alto costo de la reconstrucción del modelo y la recolección de datos de entrenamiento obstruye el desarrollo de algoritmos.



- ▶ Introducción
- ▶ Desplazamiento del conjunto de datos
- ▶ Aprendizaje de transferencia
- ▶ Adaptación de dominio



- La alta variabilidad de las señales de EEG se han estudiado como un problema del desplazamiento en el conjunto de datos (*dataset shift* en inglés).
- Sea  $(\mathbf{X}, \mathbf{Y})$  el conjuntos de datos definido en el espacio de muestras  $\mathcal{X}$  y el espacio de etiquetas  $\mathcal{Y}$ , ambos relacionados por una distribución conjunta  $P(\mathbf{X}, \mathbf{Y})$ .
- El *desplazamiento del conjunto de datos* es la discrepancia o la divergencia en las distribuciones conjuntas entre los datos de entrenamiento  $(\mathbf{X}_S, \mathbf{Y}_S)$  y prueba  $(\mathbf{X}_T, \mathbf{Y}_T)$ , es decir, cuando  $P_S(\mathbf{X}_S, \mathbf{Y}_S) \neq P_T(\mathbf{X}_T, \mathbf{Y}_T)$ , lo cual puede ser escrito como  $P_S(\mathbf{X}_S)P_S(\mathbf{Y}_S|\mathbf{X}_S) \neq P_T(\mathbf{X}_T)P_T(\mathbf{Y}_T|\mathbf{X}_T)$ .
- La presencia del desplazamiento del conjunto de datos hace difícil usar clasificadores convencionales para categorizar eficazmente las señales de EEG.



En la clasificación de señales de EEG, se estudian usualmente dos tipos de desplazamientos:

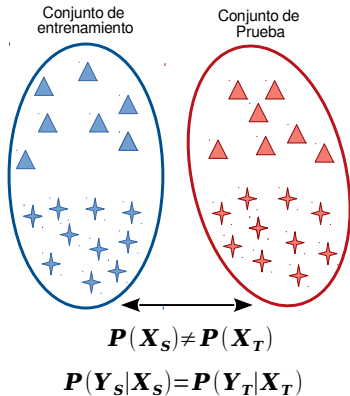
- **Desplazamiento de covariable (*Covariate shift* en inglés):** Esto sucede cuando existe un cambio en la distribución de las variables de entrada para las etapas de entrenamiento y prueba, es decir,  $P_S(\mathbf{X}_S) \neq P_T(\mathbf{X}_T)$ , pero se parte del supuesto que  $P_S(\mathbf{Y}_S|\mathbf{X}_S) = P_T(\mathbf{Y}_T|\mathbf{X}_T)$ .
- **Desplazamiento de concepto (*Concept shift* en inglés):** la relación entre las variables de entrada  $\mathbf{X}$  y las etiquetas de clase  $\mathbf{Y}$  cambia en las fases de entrenamiento y prueba, es decir,  $P_S(\mathbf{Y}_S|\mathbf{X}_S) \neq P_t(\mathbf{Y}_T|\mathbf{X}_T)$ , pero  $P_S(\mathbf{X}_S) = P_T(\mathbf{X}_T)$ .

# Tipos de desplazamiento

## 2 Desplazamiento del conjunto de datos

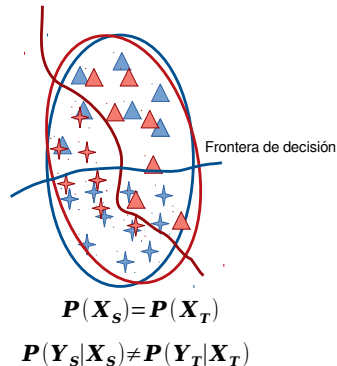


### Desplamamiento de covariable



### Desplamamiento de concepto

Frontera de decisión



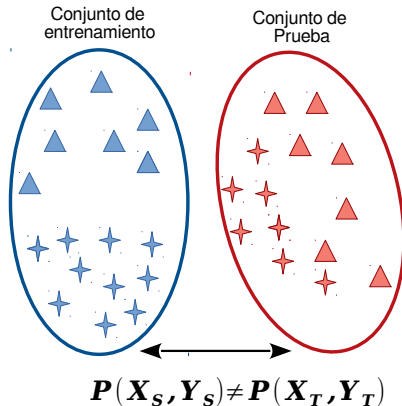


# Desplazamiento del conjunto de datos

## 2 Desplazamiento del conjunto de datos



Cambio en las distribuciones conjuntas:





- ▶ Introducción
- ▶ Desplazamiento del conjunto de datos
- ▶ **Aprendizaje de transferencia**
- ▶ Adaptación de dominio



- El aprendizaje por transferencia aborda estos problemas ajustando un modelo de clasificación basado en conocimiento previo para hacerlo adaptable en nuevas tareas.
- El aprendizaje de transferencia se enfoca en aplicar el conocimiento aprendido en un dominio conocido a uno diferente, pero relacionado.
- En otras palabras, el aprendizaje de transferencia permite a un sistema aplicar los conocimientos y habilidades aprendidas en tareas anteriores a una nueva tarea.



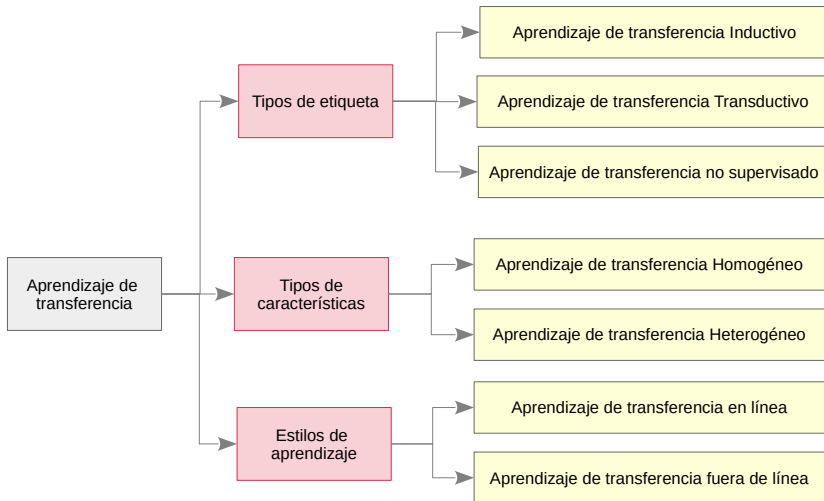
- **Dominio:** Un dominio  $\mathcal{D}$  consiste en el espacio de características  $\mathcal{X}$  de  $d$  dimensiones y la distribución de probabilidad  $P(\mathbf{X})$  de  $\mathcal{X}$ , donde  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathcal{X}$ . En aprendizaje de transferencia, el dominio que contiene el conocimiento conocido se denomina dominio fuente, representado por  $\mathcal{D}_S$ , mientras el dominio que contiene el conocimiento por aprender se denomina dominio objetivo, representado por  $\mathcal{D}_T$ .
- **Tarea:** Una tarea es el objetivo de aprendizaje, el cual consiste en un espacio de etiquetas  $\mathcal{Y}$  y la función de clasificación (escrita como  $P(\mathbf{Y}|\mathbf{X})$ , lo que significa la probabilidad de  $\mathbf{Y}$  bajo la condición  $\mathbf{X}$ ), donde  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\} \in \mathcal{Y}$ . De acuerdo a la definición de tarea, el espacio de etiquetas del dominio fuente y el dominio objetivo se representan como  $\mathcal{Y}_S$  y  $\mathcal{Y}_T$ .



- **Aprendizaje de transferencia:** Dado un dominio fuente  $\mathcal{D}_S$  y un dominio objetivo  $\mathcal{D}_T$ ;  $\mathcal{T}_S$  y  $\mathcal{T}_T$  son las tareas de  $\mathcal{D}_S$  y  $\mathcal{D}_T$ , donde  $\mathcal{D}_S \neq \mathcal{D}_T$  o  $\mathcal{T}_S \neq \mathcal{T}_T$ . El objetivo de aprendizaje de transferencia consiste en aplicar el conocimiento en  $\mathcal{D}_S$  para ayudar a aprender el conocimiento en  $\mathcal{D}_T$ , donde  $\mathcal{D}_S = \{\mathbf{X}_S, P_S(\mathbf{X}_S)\}$ ,  $\mathcal{D}_T = \{\mathbf{X}_T, P(\mathbf{X}_T)\}$ ,  $\mathcal{T}_S = \{\mathbf{Y}_S, P(\mathbf{Y}_S|\mathbf{X}_S)\}$  y  $\mathcal{T}_T = \{\mathbf{Y}_T, P(\mathbf{Y}_T|\mathbf{X}_T)\}$ .
- La idea del aprendizaje de transferencia consiste en reducir la diferencia entre dominios o tareas para asegurar un espacio de características o de etiquetas similares.

# Tipos de aprendizaje de transferencia

## 3 Aprendizaje de transferencia





- **Aprendizaje de transferencia inductivo:** las etiquetas de los dominios fuente y objetivo son conocidas, mientras que las tareas fuente y objetivo son diferentes ( $\mathcal{T}_S \neq \mathcal{T}_T$ ).
- **Aprendizaje de transferencia transductivo:** las etiquetas del dominio objetivo son desconocidas, mientras que el dominio fuente tiene una cantidad alta de datos de entrenamiento etiquetado disponible. Las tareas de los dominios fuente y objetivo son las mismas pero una diferencia en los dominios, es decir,  $\mathcal{T}_S = \mathcal{T}_T$  y  $\mathcal{D}_S \neq \mathcal{D}_T$ .
- **Aprendizaje de transferencia no supervisado:** los datos en el dominio fuente y el dominio objetivo no están etiquetados, mientras que las tareas fuente y objetivo son diferentes ( $\mathcal{T}_S \neq \mathcal{T}_T$ ).



- **Aprendizaje de transferencia homogéneo:** la semántica y las dimensiones del espacio de variables en el dominio fuente y el dominio objetivo son las mismas.
- **Aprendizaje de transferencia heterogéneo:** la semántica y las dimensiones del espacio de variables en el dominio fuente y el dominio objetivo son diferentes.





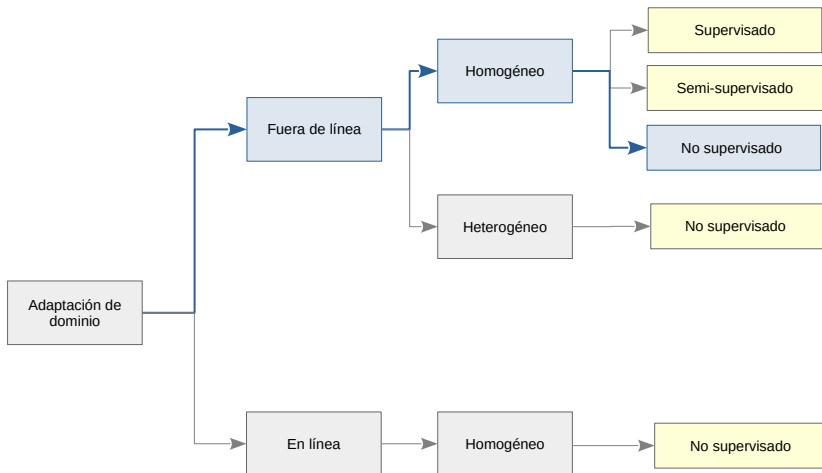
- **Aprendizaje de transferencia fuera de línea:** se realiza una sola transferencia de conocimiento para ajustar el modelo.
- **Aprendizaje de transferencia en línea:** Este tipo de aprendizaje actualiza el modelo a medida que se generan nuevos datos.



- ▶ Introducción
- ▶ Desplazamiento del conjunto de datos
- ▶ Aprendizaje de transferencia
- ▶ Adaptación de dominio



- La adaptación del dominio es un sub-campo del aprendizaje por transferencia que tiene como propósito adaptar un modelo existente a la distribución de datos de un dominio objetivo.
- La adaptación de dominio supone que hay un cambio o divergencia en las distribuciones conjuntas entre los dominios fuente y objetivo, mientras que las tareas de aprendizaje para ambos dominios son las mismas ( $\mathcal{T}_S = \mathcal{T}_T$ ).
- La adaptación de dominio supone que no existe una disponibilidad de las etiquetas en las muestras del dominio objetivo.





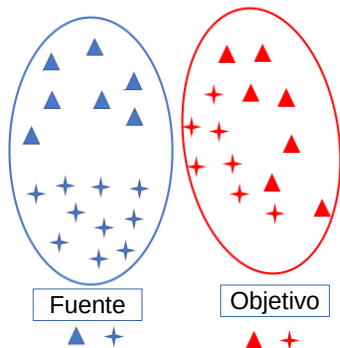
- Sea un dominio  $\mathcal{D}$  definido como el par del espacio de muestras  $\mathcal{X}$  y espacio de etiquetas  $\mathcal{Y}$ , ambos relacionados por una distribución conjunta  $P(\mathbf{X}, \mathbf{Y})$ , es decir,  $\mathcal{D} = \{\mathcal{X} \times \mathcal{Y}, P(\mathbf{X}, \mathbf{Y})\}$ .
- Sea un conjunto de datos  $(\mathbf{X}_S, \mathbf{Y}_S) = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_S}$  de  $N_S$  ejemplos etiquetados producidos de un *dominio fuente*  $\mathcal{D}_S = \{\mathcal{X} \times \mathcal{Y}, P_S(\mathbf{X}_S, \mathbf{Y}_S)\}$ , así como un conjunto de datos  $(\mathbf{X}_T) = \{\mathbf{x}_i\}_{i=1}^{N_T}$  de  $N_T$  ejemplos no etiquetados producidos de un *dominio objetivo*  $\mathcal{D}_T = \{\mathcal{X} \times \mathcal{Y}, P_T(\mathbf{X}_T, \mathbf{Y}_T)\}$ , donde  $P_S(\mathbf{X}_S, \mathbf{Y}_S) \neq P_T(\mathbf{X}_T, \mathbf{Y}_T)$ .
- La *adaptación de dominio* consiste en aprender un clasificador  $G(\cdot)$  que predice correctamente las etiquetas sobre los datos del dominio objetivo, reduciendo las diferencias de las distribuciones conjuntas entre el dominio fuente  $\mathcal{D}_s$  y el dominio objetivo  $\mathcal{D}_t$ , tal que  $P_S(\mathbf{X}_S, \mathbf{Y}_S) \approx P_T(\mathbf{X}_T, \mathbf{Y}_T)$ .

# Diferencias en las distribuciones marginal y condicional

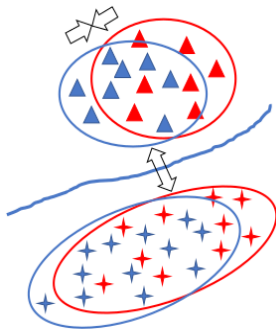
## 4 Adaptación de dominio



Antes de la adaptación



Después de la adaptación

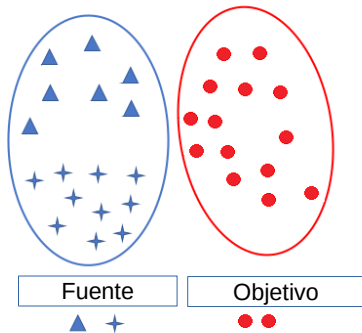


# No hay etiquetas en el dominio objetivo

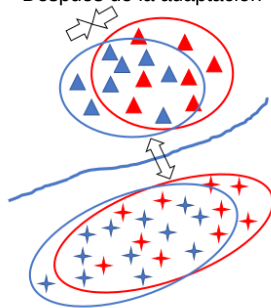
## 4 Adaptación de dominio



Antes de la adaptación



Después de la adaptación





La adaptación de dominio tiene dos objetivos competitivos:

1. Discriminabilidad: la capacidad de discriminar entre datos provenientes de diferentes clases dentro de un dominio particular.
2. Invariancia de dominio: la capacidad de medir la similitud entre las clases de datos en todos los dominios.

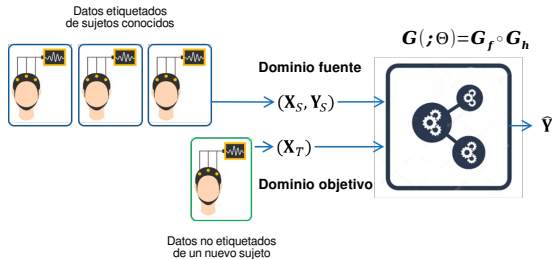




- La función de etiquetado  $G : \mathbf{X} \rightarrow \mathbf{Y}$  es una red neuronal profunda parametrizada por un conjunto de pesos  $\Theta$ .
- $G$  puede ser expresada como una composición de dos funciones,  $G_f \circ G_h$ , donde  $G_f : \mathbf{X} \rightarrow \mathbf{Z}$  es un extractor de características con pesos  $\Theta_f$ , mientras la función  $G_h : \mathbf{Z} \rightarrow \mathbf{Y}$  es un etiquetador de características con pesos  $\Theta_h$ , y  $\mathbf{Z}$  es el espacio de representación latente.
- La *adaptación de dominio profunda* pretende aprender una red neuronal profunda  $G$  que predice correctamente las etiquetas sobre los datos del dominio objetivo, obteniendo una representación latente  $\mathbf{Z}$ , donde  $P_S(\mathbf{X}_S, \mathbf{Y}_S) \approx P_T(\mathbf{X}_T, \mathbf{Y}_T)$ .

# Adaptación de dominio en la clasificación de señales de EEG





## 4 Adaptación de dominio



- Métodos de adaptación de dominio profunda ajustan los pesos  $\Theta$  de la red neuronal profunda  $G$  para reducir la discrepancia entre las distribuciones de dominio en un espacio de representación de características  $\mathbf{Z}$ , minimizando la siguiente función de costo:

$$\mathcal{L}(\mathbf{X}_S, \mathbf{Y}_S, \mathbf{X}_T, G; \Theta_f, \Theta_h) = \underbrace{\mathcal{L}_{\text{cls}}(\mathbf{X}_S, \mathbf{Y}_S, G; \Theta_f, \Theta_h)}_{\text{Función de clasificación}} + \lambda_{\text{dis}} \cdot \underbrace{\mathcal{L}_{\text{dis}}(\mathbf{X}_S, \mathbf{X}_T, G_f; \Theta_f)}_{\text{Función de discrepancia}} \quad (1)$$



-  Bashivan, P., Bidelman, G. M., and Yeasin, M. (2014).  
Spectrotemporal dynamics of the eeg during working memory encoding and maintenance predicts individual behavioral capacity.  
*European Journal of Neuroscience*, 40(12):3774–3784.
-  Bashivan, P., Rish, I., Yeasin, M., and Codella, N. (2016).  
Learning representations from eeg with deep recurrent-convolutional neural networks.  
In *International Conference on Learning Representations (ICLR) 2016*, pages 1–14.
-  Jiménez-Guarneros, M. and Gómez-Gil, P. (2017).  
Cross-subject classification of cognitive loads using a recurrent-residual deep network.  
In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–7. IEEE.
-  Wan, Z., Yang, R., Huang, M., Zeng, N., and Liu, X. (2021).  
A review on transfer learning in eeg signal analysis.  
*Neurocomputing*, 421:1–14.