Selección de candidatos informativos en datos astronómicos

Javiera Astudillo

Profesores: Karim Pichara (DCC, PUC); Pavlos Protopapas (IACS, Harvard)

Tabla de contenidos

Aprendizaje de Máquinas

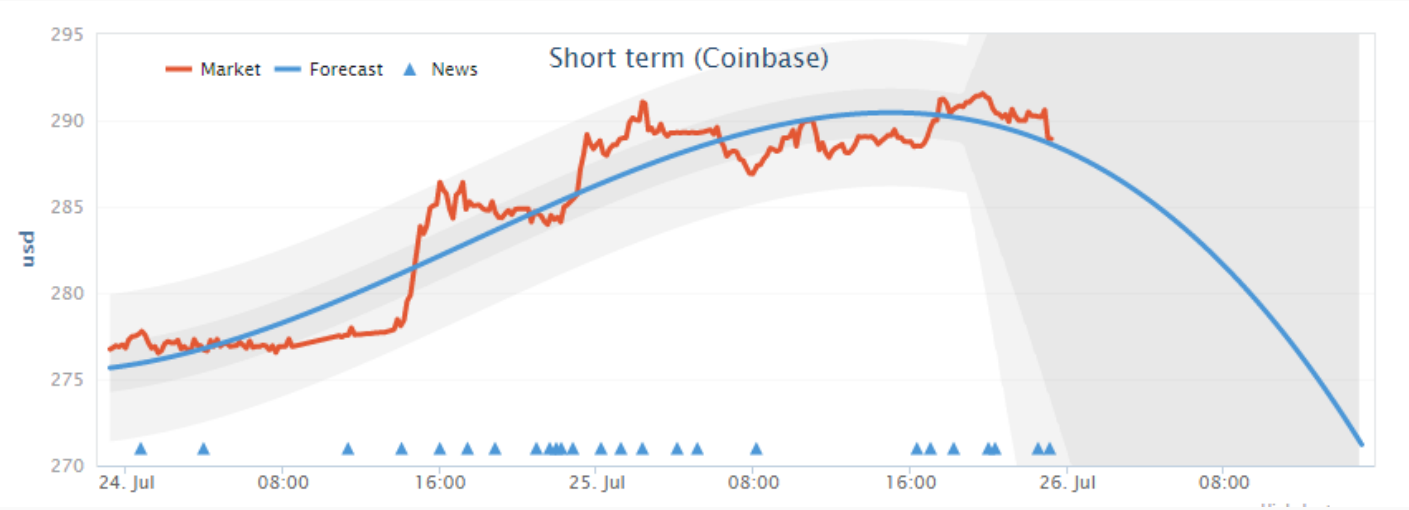
¿Qué es?

Aprender de los datos.

- Haciendo uso de herramientas de probabilidades y estadística.
- Sin indicarle qué aprender (descubrimientos de patrones no vistos).

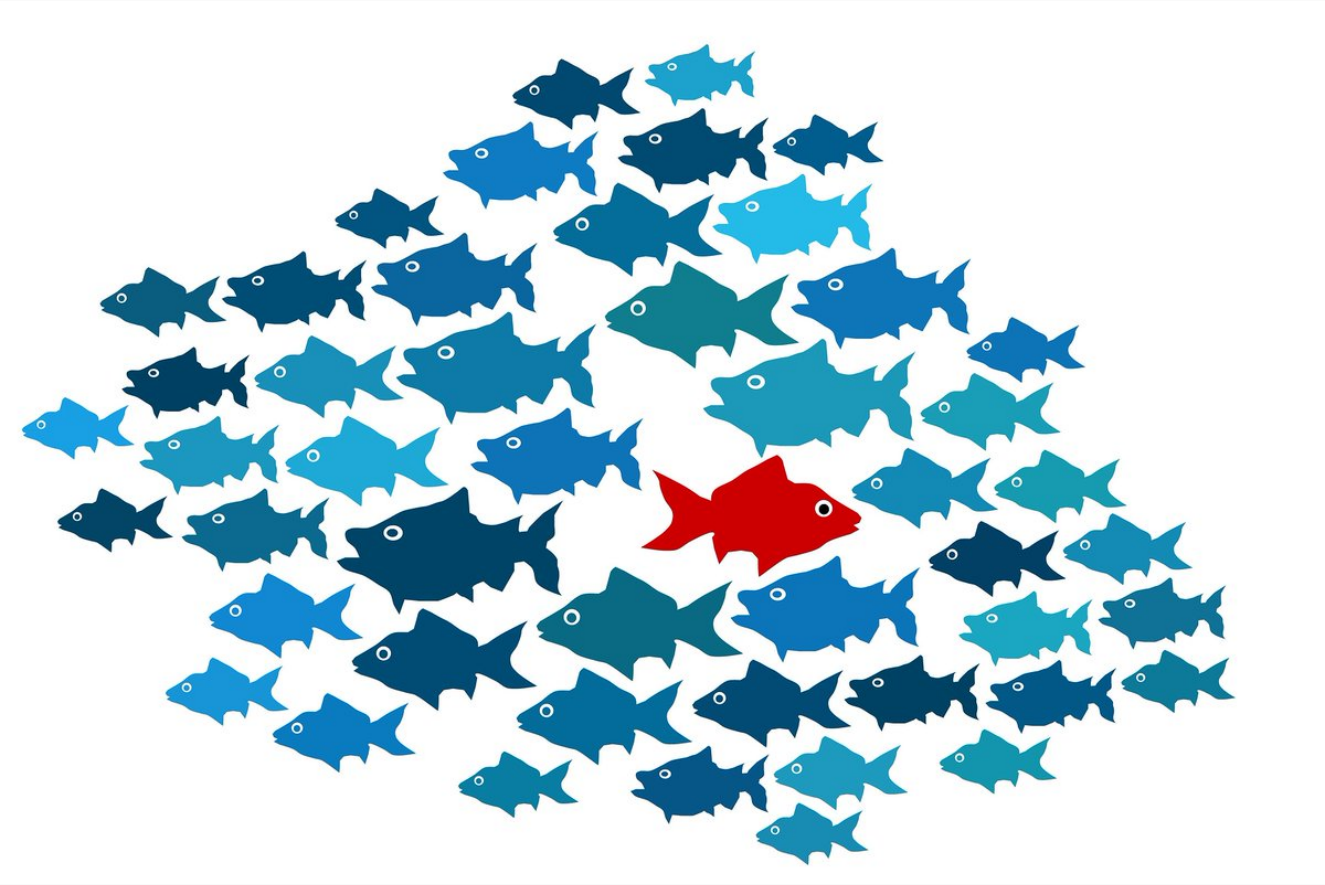
Ejemplo: clasificación

Ejemplo: regresión

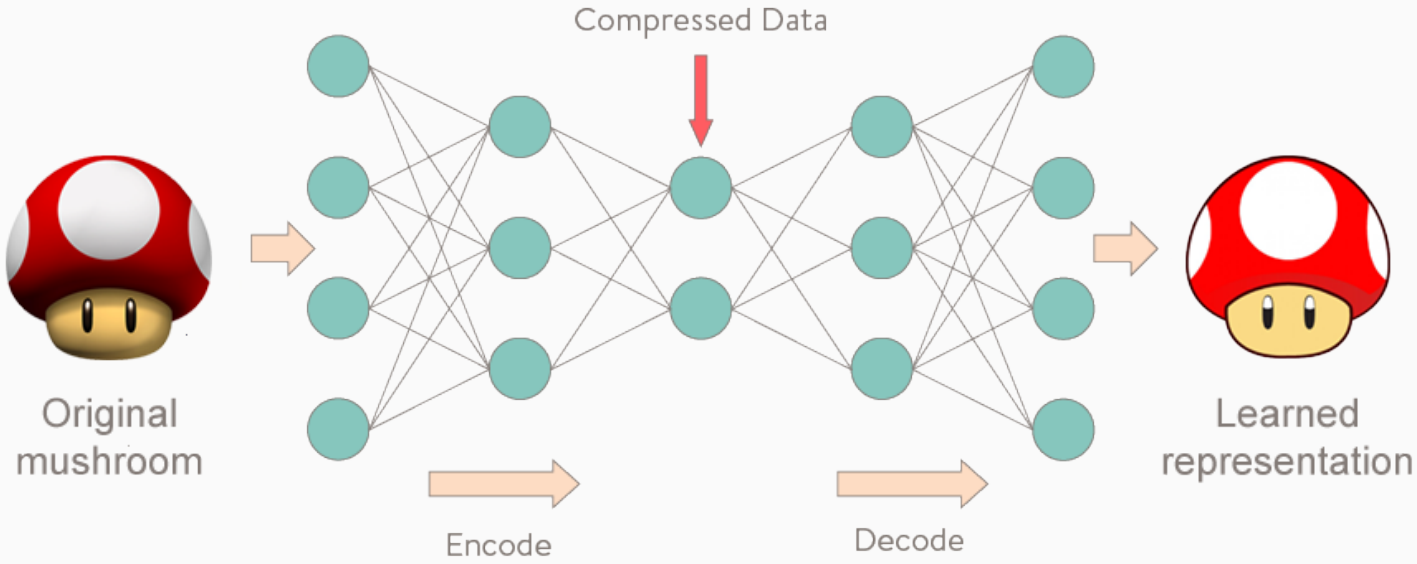


Ejemplo: clusterización

Ejemplo: detección de outliers

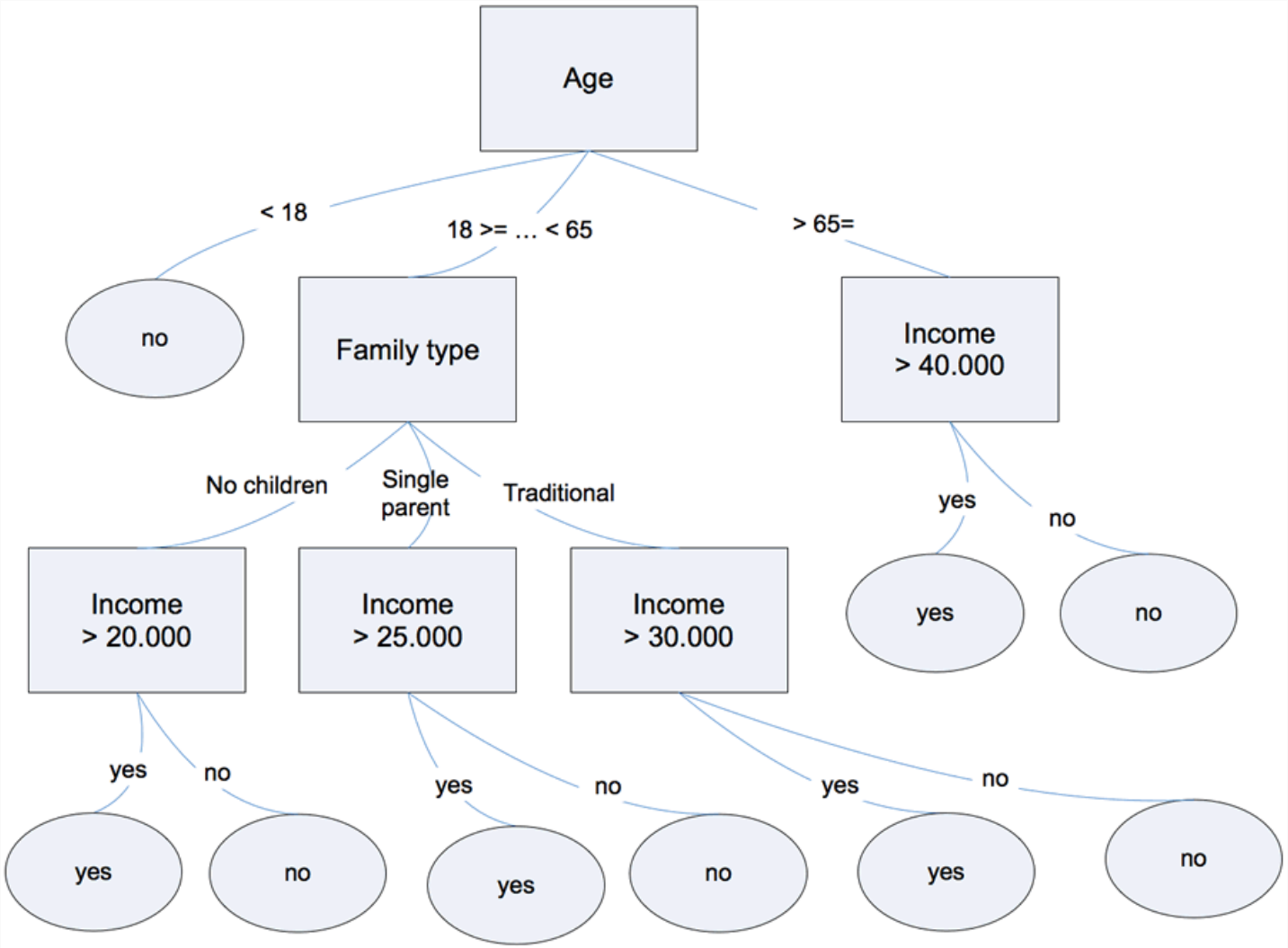


Ejemplo: reducción de dimensionalidad

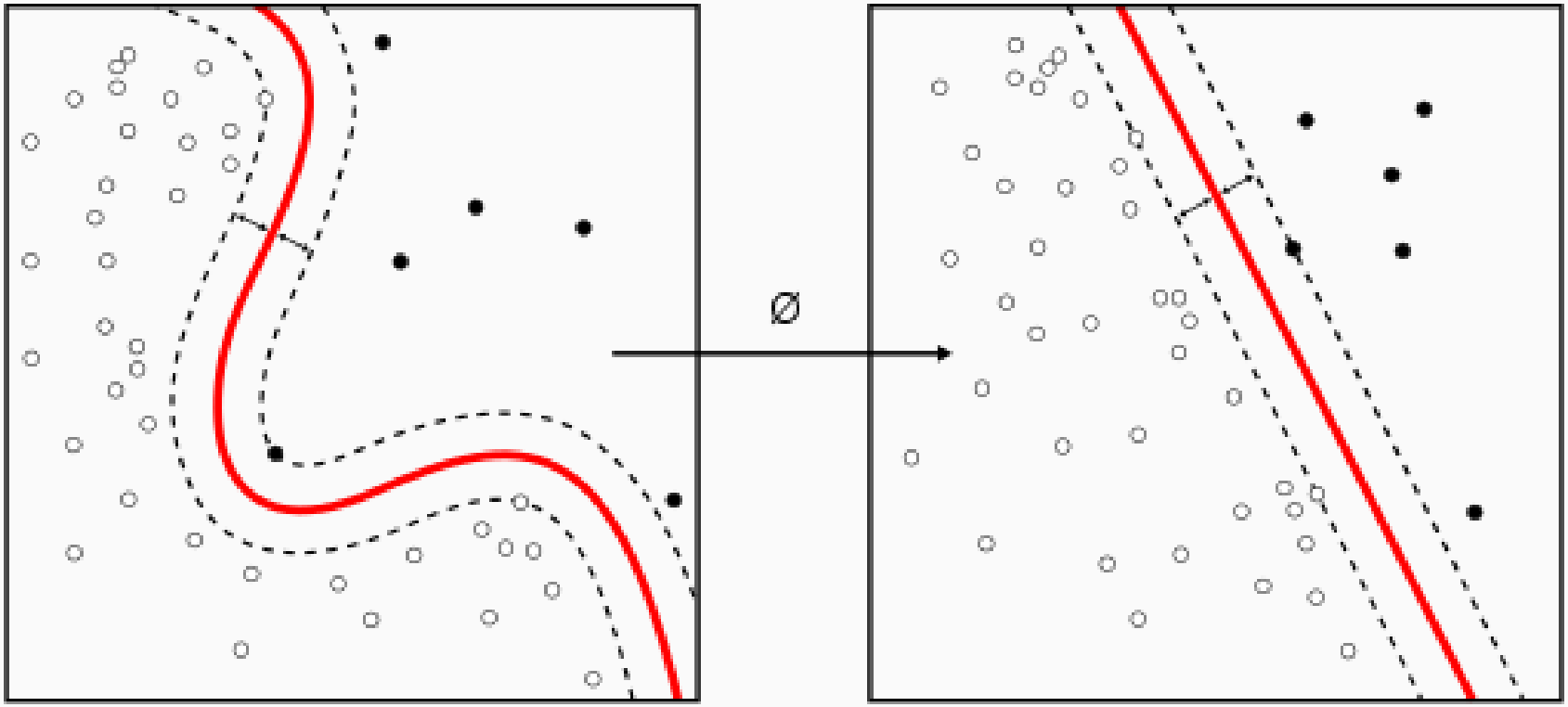


Algoritmos y técnicas (algunos)

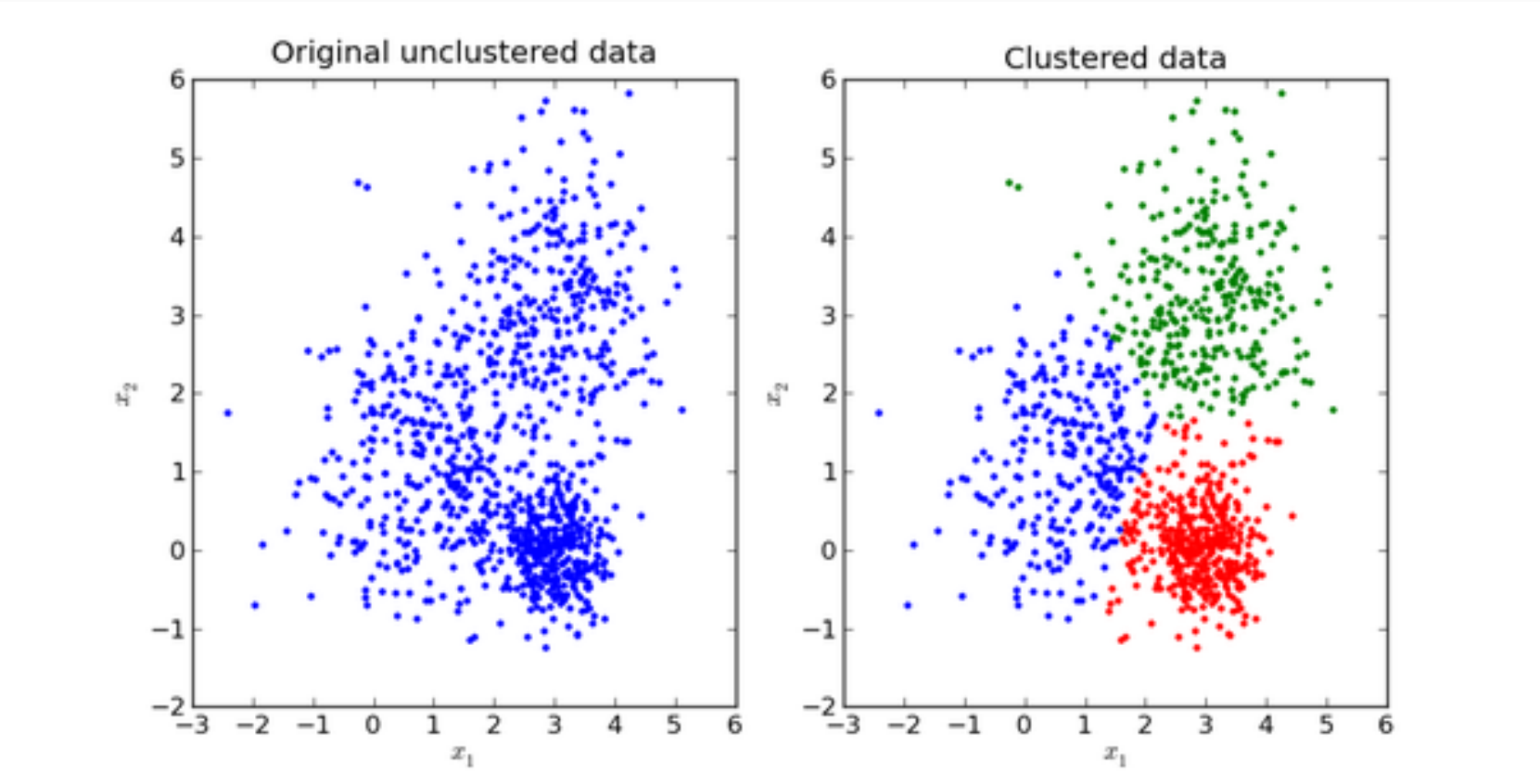
Árboles de decisión: predicción, regresión



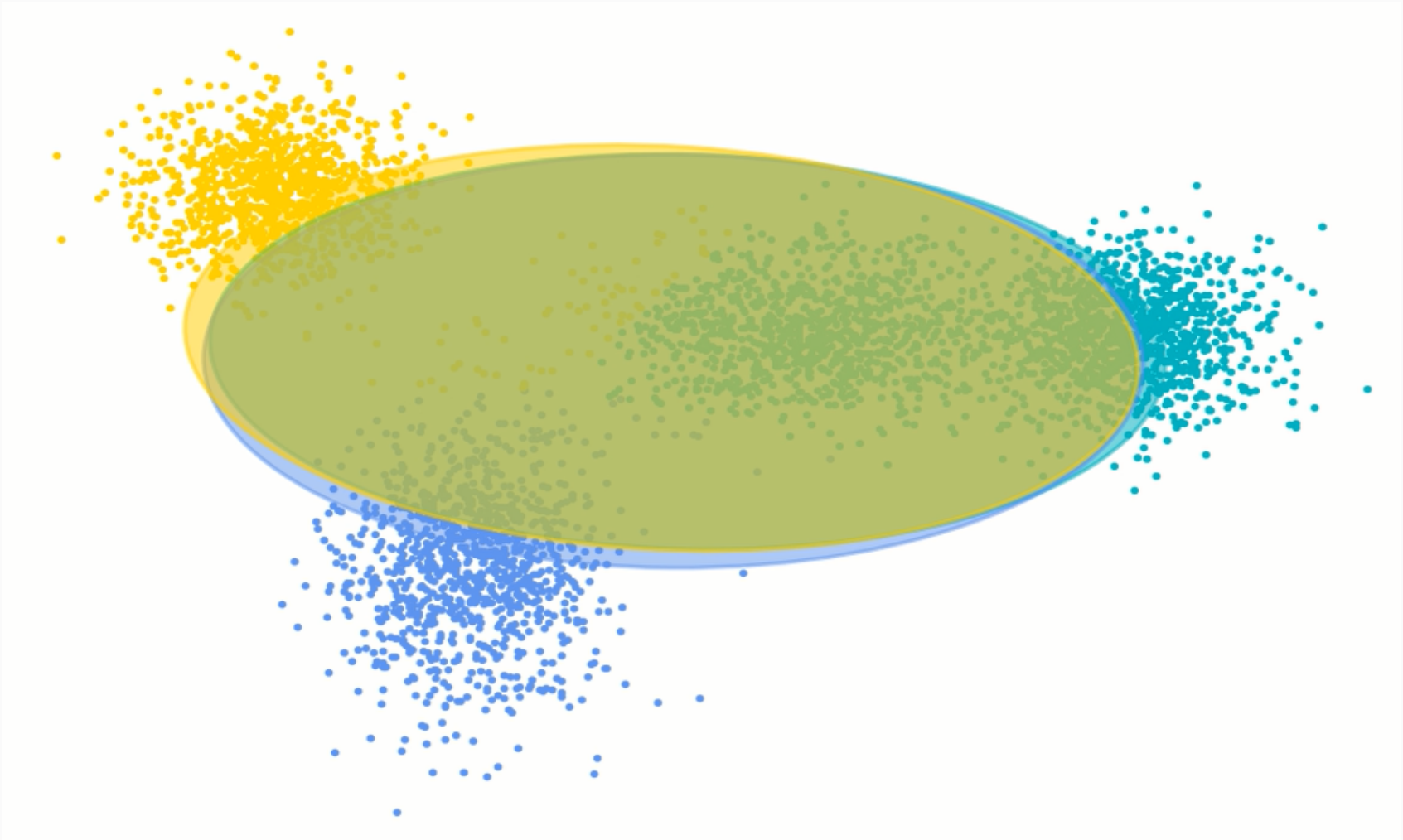
Support Vector Machine: clasificación



Mezcla de Gaussianas (GMM)

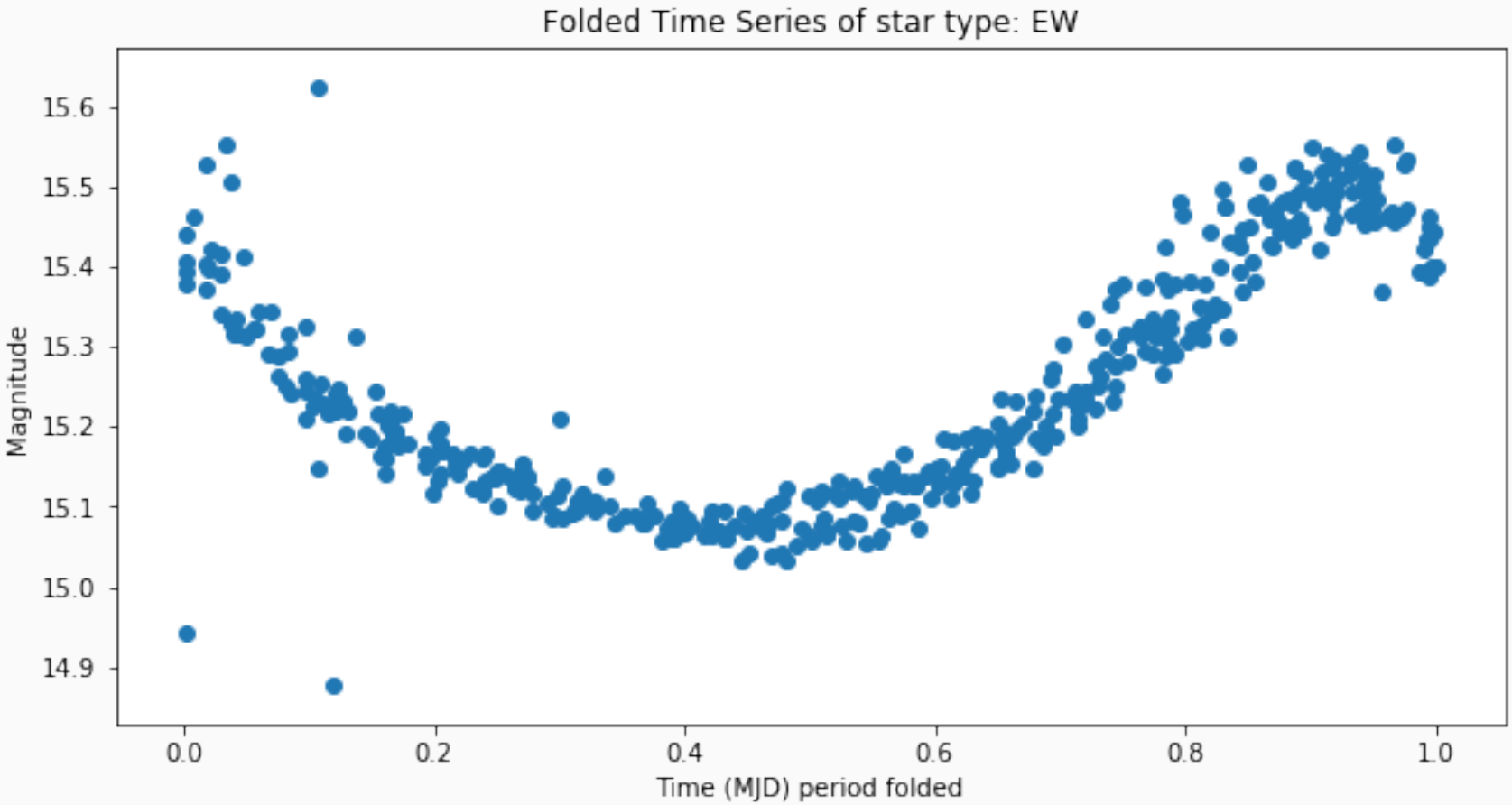


GMM (Gaussian Mixture Model)

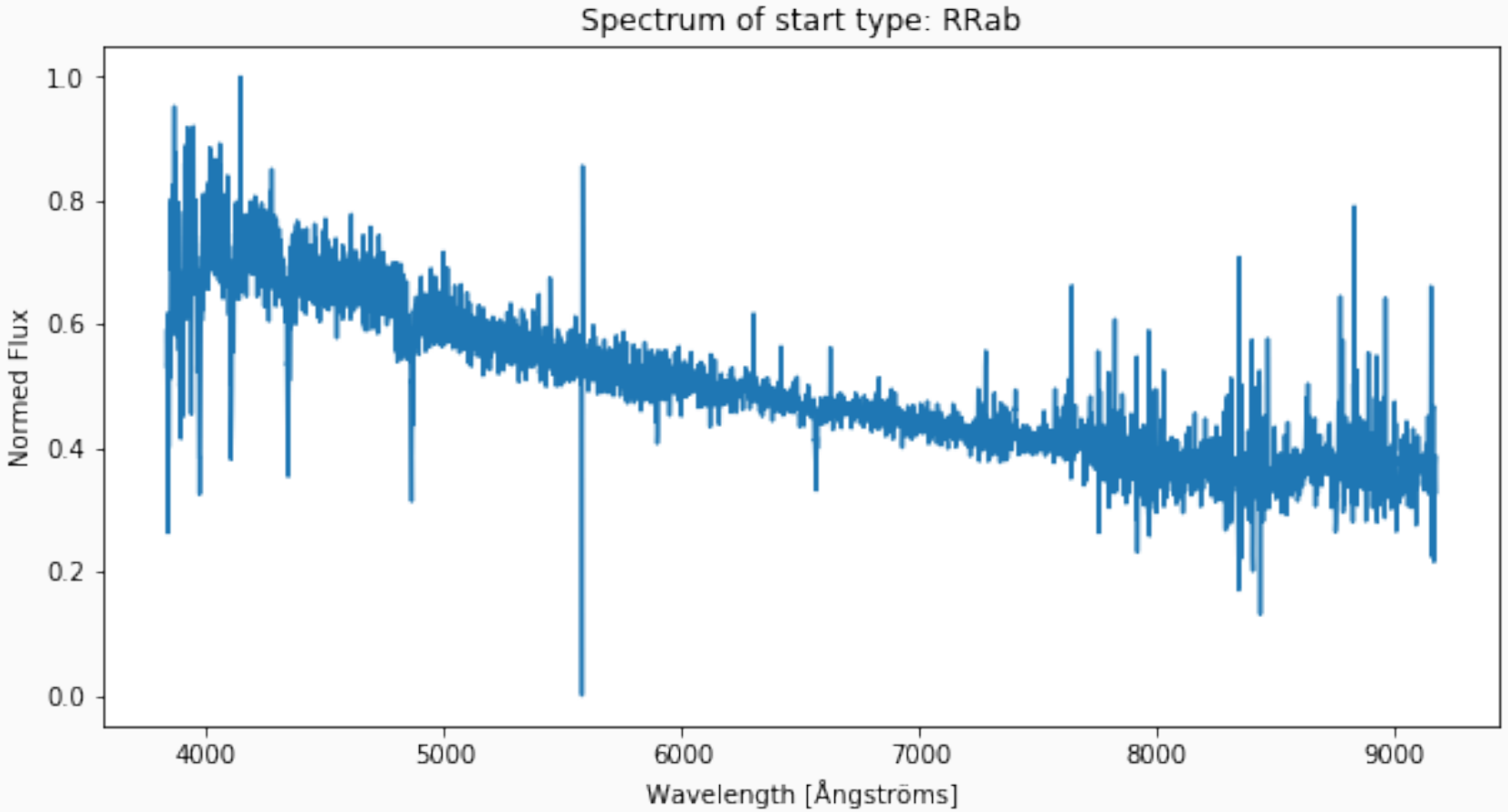


Trabajo de Investigación

Datos: Series de Tiempo



Datos: Espectros



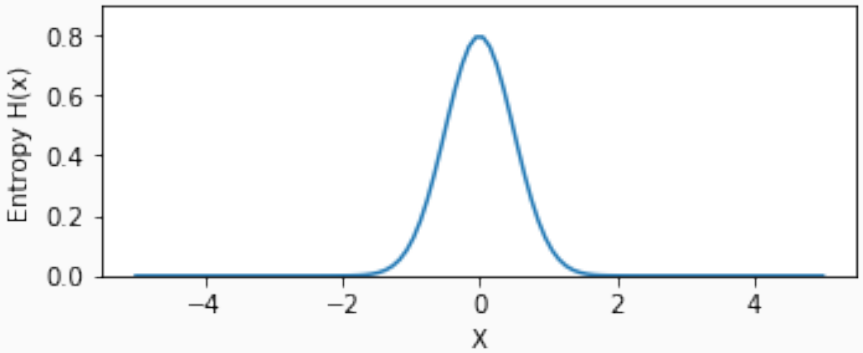
Problema

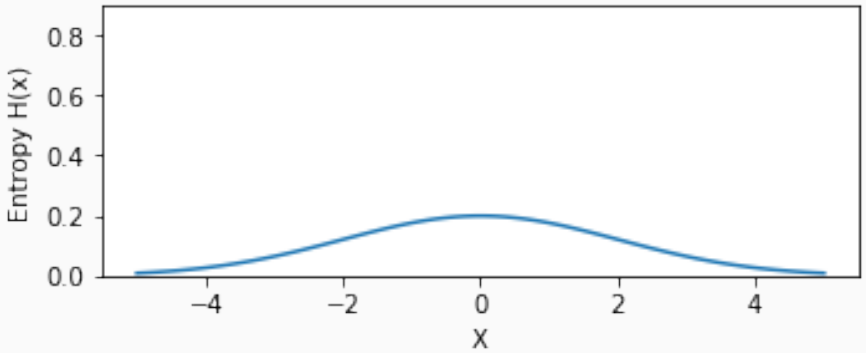
- Recursos limitados
- Seleccionar objetos para *follow-up*
- ¿Criterio?

Criterio: información

- ¿Qué es información?
  - En este trabajo: mejor clasificación.
    - Estimación del desempeño de la clasificación: Entropía

Entropía

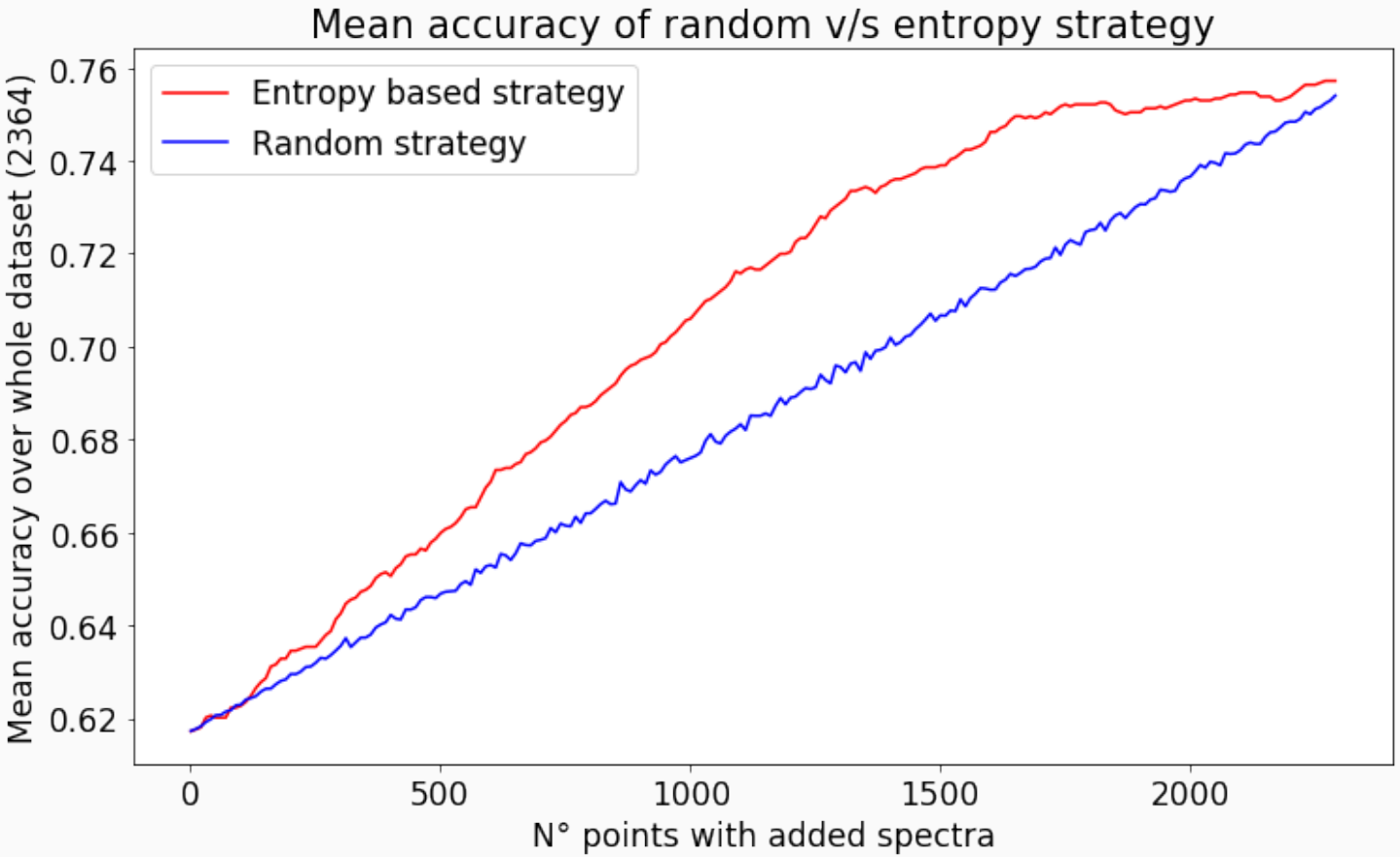

Entropy H(x)
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
X


Entropy H(x)
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
X

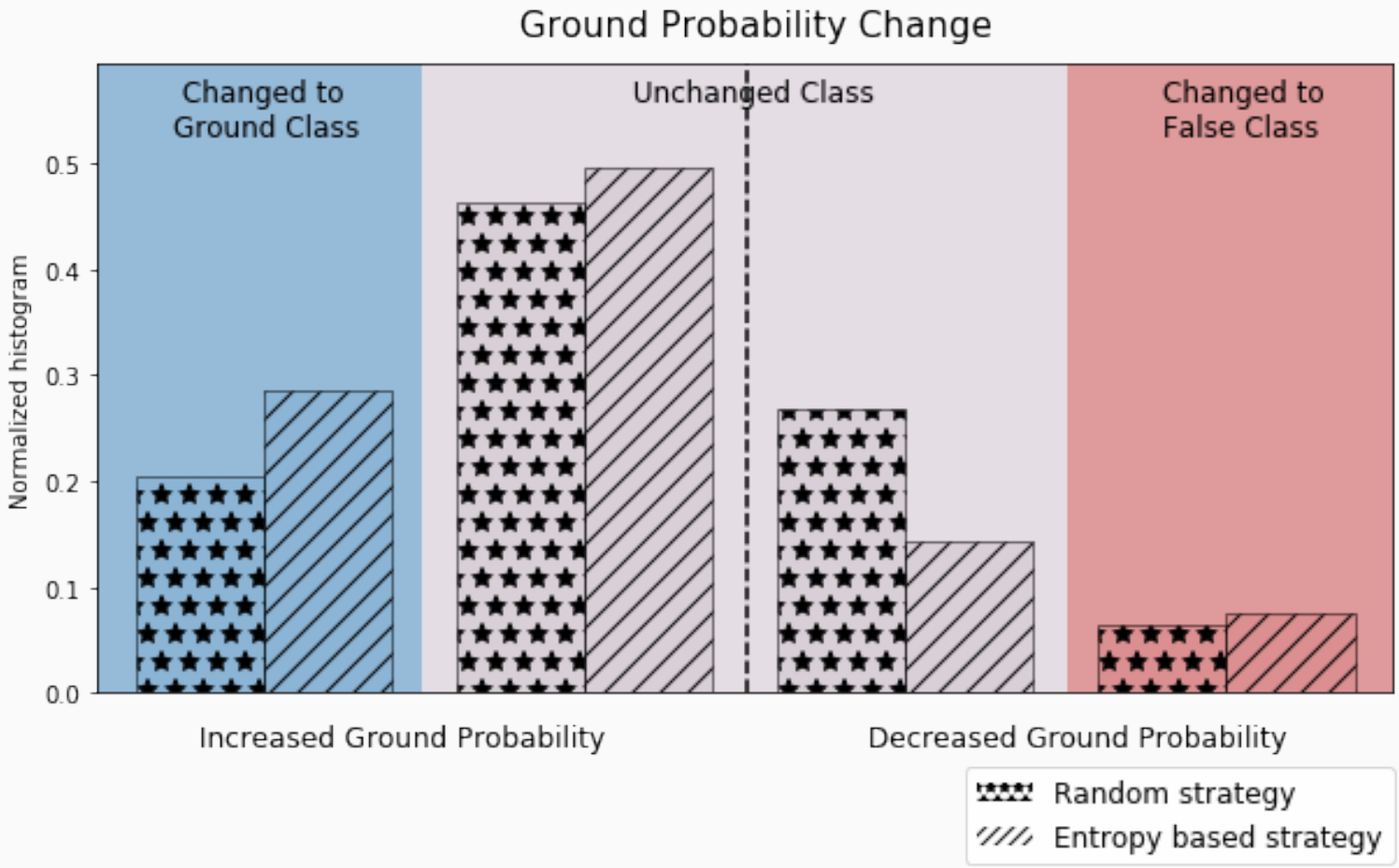
Metodología (idea)

- Extraemos características de series de tiempo y espectros a través de Auto Encoders variacionales (VAE).
- Entrenamos clasificadores solo con series de tiempo.
- Entrenamos clasificadores con series de tiempo y espectros.
- Estimamos el espectro de un objeto, dada su serie de tiempo.
- Estimamos el cambio de entropía en la clasificación con los modelos anteriores.
- Seleccionamos candidatos que presenten un gran cambio en la entropía.

Resultados


Mean accuracy of random v/s entropy strategy
Entropy based strategy
Random strategy
Mean accuracy over whole dataset (2364)
0.76
0.74
0.72
0.70
0.68
0.66
0.64
0.62
0
500
1000
1500
2000
N° points with added spectra

Resultados


Ground Probability Change
Changed to Ground Class
Unchanged Class
Changed to False Class
Normalized histogram
0.0
0.1
0.2
0.3
0.4
0.5
Increased Ground Probability
Decreased Ground Probability
Random strategy
Entropy based strategy

Conclusiones

- Primer approach a abordar este problema.
- Criterio de selección mejor que aleatorio.
- Hay una correlación entre entropía y desempeño de clasificación.
- Aún quedan varias mejoras como trabajo futuro, tales como el criterio de mejora y los diversos modelos usados.

Preguntas?

Metodología