

# Examen rápido No. 3

Reconocimiento de Patrones (2023-2)

Julio Waissman Vilanova

Nombre: \_\_\_\_\_

## Problema 1

El *ruido determinista* depende mucho de  $\mathcal{H}$ , ya que hay modelos que aproximan mejor, de forma natural, la función desconocida  $f$  que otros.

Bajo la hipótesis que  $\mathcal{H}' \subset \mathcal{H}$  para una  $f$  fija, ¿Cómo esperas que se comporte el ruido determinista si utilizamos  $\mathcal{H}'$  en lugar de  $\mathcal{H}$  como modelo de aprendizaje?:

- Esperas que el ruido determinista disminuya
- Esperas que el ruido determinista aumente
- Esperas que el ruido determinista sea el mismo
- No es razonable esperar una relación entre el ruido determinista de  $\mathcal{H}$  y  $\mathcal{H}'$ .

## Problema 2

Consideremos el problema de aprendizaje de clasificación binaria tal que  $x \in \mathbb{R}^n$ ,  $y \in \{-1, 1\}$ , en el cual separamos nuestro problema de aprendizaje en dos partes:

1. Encontrar  $\hat{a}(x) = \Pr(y = 1|x) = \sigma(w^T x + b)$ , donde  $\sigma$  es la función logística,  $w \in \mathbb{R}$  y  $b \in \mathbb{R}$  son los parámetros que se relacionan linealmente con la entrada. Esta es la etapa de regresión, y como usamos la función logística, pues le llamaremos regresión logística.
2. Encontrar  $\hat{y} = \text{sign}(\hat{a}(x) - \theta)$  donde  $\theta \in [0, 1]$  es el umbral de probabilidad mínima en la que se considera que  $x$  debe de pertenecer a la clase distinguida.
  1. Esta es la etapa de decisión.

Para este problema vamos a considerar la función de pérdida de tipo bisagra (Hinge loss), la cual está dada por:

$$loss_h(a, \hat{a}) = \max(0, 1 - a\hat{a})$$

Responde lo siguiente:

- ¿Cual es la diferencia entre la función de perdida propuesta aquí, y la vista en clase (mínimo de entropía)?
- Encuentra como calcular  $\Delta w$  y  $\Delta b$  tales que, para el método de aprendizaje por descenso de gradiente cada *epoch* actualice los parámetros del modelo de la forma siguiente:

$$w \leftarrow w + \eta \Delta w, \quad b \leftarrow w + \eta \Delta w,$$

donde  $\eta \in \mathbb{R}$  es la tasa de aprendizaje.