

Examen rápido No. 2

Reconocimiento de Patrones (2023-2)

Julio Weissman Vilanova

Nombre: _____

Problema 1

Consideremos una función conocida $f : [-1, 1] \rightarrow \mathbb{R}$ dada por $f(x) = \sin(\pi x)$ y con una muestra de datos de entrenamiento tomada a partir de una distribución uniforme. Vamos a asumir que el conjunto de aprendizaje sólo tiene dos datos tomados independientemente, y que el algoritmo de aprendizaje encuentra el valor que minimiza el error cuadrado medio (MSE) en los ejemplos.

Si consideramos una hipótesis del tipo:

$$h(x) = ax$$

- ¿Cuál es la fórmula para obtener el valor óptimo a^* con dos valores $\mathcal{D} = \{(x_1, y_1), (x_2, y_2)\}$?

- ¿Cuál sería el valor de

$$\bar{h}(x) = E_{\mathcal{D} \in \{\mathcal{D}_1, \dots, \mathcal{D}_k\}}[h^{\mathcal{D}}]$$

si se pone en función de x (i.e. $\bar{h}(x) = \bar{a}x$):

Problema 2

Para el modelo considerado en el problema 1, encierra la hipótesis que crees que tendría el menor error fuera de muestra:

1. $h(x) = b$
2. $h(x) = ax$
3. $h(x) = ax + b$
4. $h(x) = ax^2$
5. $h(x) = ax^2 + b$

- Explica tus razones:

Problema 3

Asumimos que tenemos un conjunto de K posibles modelos a usar (o su combinación usando alguna técnica de *AutoML*) $\mathcal{H}_1, \dots, \mathcal{H}_K$. Asumimos también que todas las hipótesis tienen una dimensión VC finita.

- ¿Cuales de las siguientes aseveraciones son correctas, si decidimos usar todos los modelos (esto es, la unión de todos los modelos)?:

1. $0 \leq d_{vc}(\bigcup_{k=1}^K \mathcal{H}_k) \leq \sum_{k=1}^K d_{vc}(\mathcal{H}_k)$
2. $0 \leq d_{vc}(\bigcup_{k=1}^K \mathcal{H}_k) \leq K - 1 + \sum_{k=1}^K d_{vc}(\mathcal{H}_k)$
3. $\min_{k \in \{1, \dots, K\}} (d_{vc}(\mathcal{H}_k)) \leq d_{vc}(\bigcup_{k=1}^K \mathcal{H}_k) \leq \sum_{k=1}^K d_{vc}(\mathcal{H}_k)$
4. $\max_{k \in \{1, \dots, K\}} (d_{vc}(\mathcal{H}_k)) \leq d_{vc}(\bigcup_{k=1}^K \mathcal{H}_k) \leq \sum_{k=1}^K d_{vc}(\mathcal{H}_k)$
5. $\max_{k \in \{1, \dots, K\}} (d_{vc}(\mathcal{H}_k)) \leq d_{vc}(\bigcup_{k=1}^K \mathcal{H}_k) \leq K - 1 + \sum_{k=1}^K d_{vc}(\mathcal{H}_k)$

- ¿Cual es la cota más ajustada entre las que son correctas?