



Interrogación 3

Pregunta 1

- a) ¿Como se podría realizar clasificación con un árbol de decisión, si falta el valor de alguna de las dimensiones del vector de entrada? **(1 pto.)**
- b) Explique por qué la cantidad de parámetros activos (distintos de cero) en una red neuronal, puede dar una noción de la complejidad del modelo. **(1 pto.)**
- c) Indique como utilizaría validación cruzada para estimar la mejor profundidad de un árbol de decisión. **(1 pto.)**
- d) ¿De qué manera el método del momentum disminuye el riesgo de caer en mínimos locales en redes neuronales con capas ocultas? **(1 pto.)**
- e) ¿En qué situaciones es preferible utilizar el radio de ganancia por sobre la ganancia de información? **(1 pto.)**
- f) ¿Como podría utilizarse una red neuronal para aumentar la resolución de imágenes (manteniendo una buena calidad)? **Hint:** enfóquese en el entrenamiento de una red para esta tarea. **(1 pto.)**

Pregunta 2

- a) Considere un problema de clasificación sobre variables categóricas. Extienda el algoritmo de construcción de los árboles de decisión basado en la ganancia de información, para que se puedan realizar tests sobre dos variables de manera simultánea **(2 ptos.)**.
- b) En general, al momento de decidir el valor a testear en un nodo de un árbol de decisión, se toma la ganancia de información como métrica. Una de las desventajas de esta, es que no considera la nueva estructura del árbol en el cálculo (la resultante de seleccionar ese test, con distinta profundidad y número de nodos), lo que puede derivar en problemas de sobreentrenamiento. Extienda la métrica de la ganancia de información, agregando un nuevo término aditivo, de manera que ahora, para tomar la decisión, se considere información sobre la posible nueva estructura del árbol. **Hint:** considere la decisión en un nodo como la minimización del riesgo estructural empírico. **(2 ptos.)**
- c) Considere una competencia, donde se debe resolver un problema de regresión en base a variables categóricas. Dado que en este problema el riesgo de sobreentrenamiento es alto, sólo se permite utilizar árboles de regresión sobre una (1) de las variables disponibles, con el fin de limitar la profundidad del árbol. Utilizando múltiples árboles de **manera secuencial** (cada árbol puede utilizar la variable que quiera), indique como es posible construir un sistema de regresión que estime de mejor manera la función buscada. **(2 ptos.)**

Pregunta 3

Dado un conjunto de ejemplos $x = \bigcup_i^N x_i$, con $x_i \in \mathbb{R}^L$ y sus etiquetas asociadas $y = \bigcup_i^N y_i$, con $y_i \in [0, 1]$, considere la función de pérdida *cross-entropy* definida a continuación:

$$E(x, y; w) = - \sum_i y_i \ln o_i + (1 - y_i) \ln(1 - o_i)$$

donde o_i es la salida de la red (perceptrón sigmoidal) para el ejemplo i y w es el vector de parámetros de la red. En base a esto, conteste las siguientes preguntas:

- a) Construya el grafo de cómputo para una red con una capa oculta de 2 neuronas con función de activación ReLU, que realice la clasificación de x , utilizando como pérdida la función $E(x, y; w)$. No combine múltiples operaciones en un sólo nodo del grafo. **(3 ptos.)**
- b) Extienda la definición de $E(x, y; w)$ para el escenario de clasificación multiclase (más de dos categorías), *i.e.*, $y_i \in [0, K]$. **(3 ptos.)**