

# I302 - Aprendizaje Automático y Aprendizaje Profundo

1<sup>er</sup> Semestre 2024

Parcial 1

- 
1. (1 punto) El intervalo de tiempo entre arribos de autos a un peaje se puede modelar como un proceso estocástico Poisson donde el tiempo entre arribos es una variable aleatoria  $T$  con una distribución de probabilidad exponencial, como se muestra abajo. Si tenemos un conjunto de muestras  $\{t_1, t_2, \dots, t_N\}$  de los intervalos de tiempo entre arribos de autos sucesivos, derive el estimador de máxima verosimilitud para el parámetro  $\beta$ .

$$P(T = t) = \beta e^{-\beta t}$$

2. (1 punto) Con respecto a un modelo Random Forest, marque cuales de las siguientes afirmaciones son verdaderas y cuales son falsas.
  - (a) Al aumentar el número de árboles se reduce la capacidad predictiva del modelo.
  - (b) Bagging elige un subconjunto aleatorio de features sobre el cual se determina la partición de un nodo (node splitting).
  - (c) Bagging busca aumentar la correlación entre árboles.
  - (d) La profundidad de los árboles se limita principalmente para reducir el costo computacional de entrenamiento.
3. (0.5 puntos) Se tiene una red neuronal densamente conectada con una capa oculta de 7 nodos, una capa de salida con 3 nodos y un vector de entradas de dimensión 4. ¿Cuántos parámetros a ajustar tiene la red?
4. (0.5 puntos) Sea una red neuronal densamente conectada con funciones de activación ReLU, aplicada a un problema de regresión con función de costo cuadrática. ¿La función de costo es convexa con respecto a los pesos de la red? Justifique su respuesta.
5. (1 punto) Marque cuales de las siguientes afirmaciones son verdaderas y cuales son falsas.
  - (a) Al aplicar SGD es necesario aleatorizar el orden en que se usan los datos de entrenamiento (de no ser así el algoritmo no converge).

- (b) Batch GD y SGD siempre convergen al mismo punto, pero uno lo hace mas rápido que el otro.
  - (c) Mini-batch GD busca reducir el tiempo de computo con respecto a Batch GD, y al mismo tiempo retener las cualidades de SGD.
  - (d) El algoritmo ADAM garantiza que se alcanzará el óptimo global.
6. (1.5 puntos) Suponga que está intentando resolver un problema de clasificación binaria donde la clase positiva se encuentra subrepresentada (por ejemplo, por cada positivo hay 9 negativos).
- (a) Describa las diferentes técnicas que podría usar durante el entrenamiento para mitigar el problema de desbalance de clases, así como los pros y contras de cada una.
  - (b) ¿Aplicaría estas técnicas en el momento de evaluar la capacidad predictiva sobre el conjunto de test? Justifique su respuesta.
7. (1.5 punto) Sea un problema de regresión para el cual se está desarrollando un modelo de red neuronal densamente conectada con funciones de activación ReLu y función de costo que es la suma de errores absolutos (en lugar de la suma de los errores cuadráticos). Derive el algoritmo de backpropagation para calcular el gradiente de la función de costo con respecto a los pesos de la red.
8. (0.5 puntos) Sea un problema de clasificación binaria para el cual se entrenó un modelo de regresión logística que utiliza un umbral de 0.5 para clasificar, cuyo vector de pesos es  $W$  y el accuracy de test es  $A$ . ¿Cambiará  $A$  si usamos pesos  $W' = W/2$ ?
9. (2 puntos) Sea un árbol de decisión para un problema de clasificación. Demostrar que el índice de Gini de un nodo alcanza el mínimo valor posible cuando el nodo es puro (es decir, la region correspondiente al nodo contiene únicamente muestras de una clase).
10. (0.5 puntos) El gráfico de abajo muestra como cambia el error de entrenamiento y el error de validación de un modelo de regresión con regularización L2, para distintos valores del hiperparámetro de regularización  $\lambda$ . ¿Qué valor de  $\lambda$  elegiría? Justifique su respuesta.

