

Nombres: Daniela Navas

Carné: 211000

HOJA DE TRABAJO 1

Task 1 – Preguntas Teóricas– Agentes Inteligentes

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. Describa la abstracción general del concepto de Agentes Inteligentes

Los agentes inteligentes son entidades definidas que perciben su entorno a través de sensores y actúan sobre él mediante actuadores. Tienen como principal característica su capacidad para tomar decisiones autónomas basadas en la información que reciben, lo que les permite adaptarse a diferentes situaciones y aprender de sus experiencias. Estos agentes pueden ser tanto físicos, como robots, virtuales, programas de software, entre otros. La inteligencia de un agente se mide por su capacidad para resolver problemas, planificar acciones y realizar tareas complejas en entornos dinámicos e inciertos.

2. Defina, tomando como base un ejemplo, la arquitectura descrita para Agentes Inteligentes

La arquitectura de un agente inteligente se puede definir como la estructura o definición que guía su funcionamiento. Pueden a ver ciertas variaciones en la arquitectura de un agente inteligente, pero en bases generales incluyen:

- **Unidad de Sensores:** Puede estar conformado por un sensor o una red de sensores, capta la información del entorno.
- **Capa de razonamiento:** Procesa la información recibida en algo entendible para el sistema, lo “traduce” a su lenguaje.
- **Modelo Interno:** Define el conocimiento del entorno o del sistema, que puede ser estático o dinámico.
- **Motor de Decisión:** Analiza la información y genera acciones. Esto puede incluir reglas programadas, algoritmos de búsqueda o técnicas de aprendizaje.
- **Actuadores:** Ejecutan las decisiones del agente en el entorno.

Como ejemplo, podemos tomar en cuenta un sistema de riego automático para jardines inteligentes que utiliza una arquitectura organizada en varias capas. La Unidad de Sensores está formada por dispositivos como sensores de humedad del suelo, temperatura ambiental y lluvia, que capturan información del entorno en tiempo real. Esta información es procesada por la Capa de Razonamiento, que traduce las lecturas en datos útiles, como identificar si el suelo está seco o si está lloviendo. El Modelo Interno contiene reglas estáticas, como los niveles mínimos de humedad aceptables para cada tipo de planta, y reglas dinámicas, como patrones climáticos locales o historial de riegos previos. Con estos datos, el Motor de Decisión analiza las condiciones y determina las acciones necesarias, como activar o detener el riego. Las decisiones se ejecutan a través de los Actuadores, que incluyen válvulas automáticas que controlan el flujo de agua.

3. De los tipos de ambientes vistos en clase, ¿cuál es el que más llamó su atención? ¿Por qué? ¿De qué forma cree que se puede aplicar en la vida real?

El ambiente que más llamó mi atención es el **ambiente parcialmente observable y el estocástico**. Ya que este tipo de entorno refleja la complejidad e incertidumbre del mundo real. En un ambiente parcialmente observable, los agentes no tienen acceso a toda la información relevante del entorno, lo que los obliga a realizar inferencias o estimaciones basadas en los datos disponibles, ya conocía el termino de parcialmente observables aplicado a un sistema de control y me parece interesante como con cierta información se puede estimar la dinámica interna del sistema. Por su parte, un ambiente estocástico introduce elementos de aleatoriedad, donde las mismas acciones no siempre producen los mismos resultados debido a factores impredecibles.

Un ejemplo claro de este tipo de ambiente es el de un sistema médico de diagnóstico asistido por inteligencia artificial. El sistema trabaja con información limitada y debe manejar incertidumbre debido a la variabilidad en cómo las enfermedades se manifiestan en diferentes individuos. A pesar de estas limitaciones, el agente puede utilizar modelos probabilísticos para proponer diagnósticos o planes de tratamiento óptimos.

4. ¿Qué concepto de los vistos en clase llamó más su atención? ¿Por qué? Defina dicho concepto

El concepto que más llamó mi atención es el de racionalidad en los agentes inteligentes, se refiere a la capacidad de un agente para tomar decisiones óptimas basadas en el conocimiento disponible y las condiciones del entorno. Un agente es considerado racional si, para cada estado del ambiente, su función de transformación mapea las acciones correctas que maximizarán su desempeño o lograrán sus objetivos. Por ejemplo, un agente que juega ajedrez evaluará diferentes movimientos posibles en función de su conocimiento del juego y las posiciones actuales de las piezas. La racionalidad implica que el agente seleccionará la acción que maximice sus probabilidades de ganar, considerando no solo su posición actual, sino también las posibles respuestas del oponente.

Task 2- Pregunta Teóricas- Manejo de Datasets y Métricas de Desempeño

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible

1. Usted está trabajando con un modelo de clasificación, usando un dataset que está desbalanceado para una de las clases

a. ¿Se podría usar este dataset sin ningún tipo de preprocesamiento? ¿Por qué?

Sí, se podría usar el dataset sin preprocesamiento, pero no es recomendable. Un dataset desbalanceado puede llevar a que el modelo de clasificación tenga un sesgo hacia la clase mayoritaria, resultando en un desempeño pobre para la clase minoritaria. Esto se debe a que el modelo puede aprender a predecir siempre la clase mayoritaria para minimizar el error, ignorando la clase minoritaria.

b. Asumiendo que sí necesita preprocesamiento, ¿qué tipo de trabajo haría al dataset?

Para balancear el dataset ya existen varias técnicas de preprocesamiento definidas, algunas de ellas son:

- **Submuestreo de la clase mayoritaria:** Reducir el número de muestras de la clase mayoritaria.

- **Sobremuestreo de la clase minoritaria:** Aumentar el número de muestras de la clase minoritaria, por ejemplo, mediante técnicas como SMOTE (Synthetic Minority Over-sampling Technique).
 - **Generación de datos sintéticos:** Crear nuevas muestras sintéticas para la clase minoritaria.
 - **Ajuste de pesos:** Asignar pesos mayores a las muestras de la clase minoritaria durante el entrenamiento del modelo.
- c. **Asumiendo que no hace ningún preprocesamiento, ¿qué métrica usaría para medir el desempeño de su modelo?**

Una métrica adecuada para evaluar el desempeño del modelo si no se realiza ningún preprocesamiento, podría ser el F1-score que considera tanto la precisión como el recall. Esto es especialmente útil en situaciones de desbalanceo, ya que proporciona una mejor medida del rendimiento en la clase minoritaria.

2. En clase solamente vimos algunas de las métricas que se pueden usar para medir el desempeño de modelos de clasificación. Investigue otras dos métricas (que no sean de las vistas en clase), asegúrese de incluir

Matthews Correlation Coefficient (MCC)

a. Definición de la métrica

El Coeficiente de Correlación de Matthews es una métrica que evalúa la calidad de las clasificaciones binarias. Tiene en cuenta verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, proporcionando un valor que varía entre -1 y +1. Un valor de +1 indica una predicción perfecta, 0 indica una predicción aleatoria y -1 indica una predicción completamente errónea.

b. Formulación matemática de la misma

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Donde TP representa Verdaderos Positivos, TN los Verdaderos Negativos, FP los Falsos Positivos y FN = Falsos Negativos.

Cohen's Kappa

a. Definición de la métrica

Cohen's Kappa es una métrica que mide la concordancia entre dos clasificadores o entre un clasificador y las etiquetas verdaderas, ajustando por la posibilidad de que las coincidencias ocurran por azar. Se utiliza para evaluar la precisión de las clasificaciones en problemas de clasificación binaria o multiclase.

b. Formulación matemática de la misma

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Donde P_o es la proporción de acuerdos observados (exactitud) y P_e es la proporción de acuerdos esperados por azar.

3. Explique de forma clara y concisa la diferencia entre RSE, MAE y MAPE como métricas de desempeño para modelos de regresión

RSE (Root Squared Error) es la raíz cuadrada del error cuadrático medio (MSE). Proporciona una medida en las mismas unidades que los datos originales y **penaliza más fuertemente los errores grandes**. **MAE** (Mean Absolute Error) es el promedio de los errores absolutos entre las predicciones y los valores reales. Es **menos sensible a los valores atípicos** en comparación con RSE. **MAPE** (Mean Absolute Percentage Error) mide el error absoluto en **términos porcentuales**, lo que permite comparar el rendimiento del modelo en diferentes escalas. Es útil cuando se requiere una **interpretación relativa** del error.

4. Defina brevemente en qué consiste un modelo de clustering y mencione una métrica (que no sea la vista en clase) para medir el desempeño de este tipo de modelos, para esta última, asegúrese de incluir

Un modelo de clustering consiste en agrupar un conjunto de objetos en subconjuntos (clusters), donde los objetos dentro de un mismo clúster son más similares entre sí que aquellos en diferentes clústeres. Es una técnica de aprendizaje no supervisado

El Dunn Index

a. Definición de la métrica

El Dunn Index es una métrica que busca identificar clústeres bien separados y compactos. Se define como la relación entre la mínima distancia entre clústeres y la máxima distancia dentro de un clúster. Un valor más alto del Dunn Index indica una mejor separación entre los clústeres y una mayor compactación dentro de cada clúster.