Quiz de Puntos Extra - Semana 3

Machine Learning (NRC-200)

Estudiante: Alexander Oviedo Fadul
Fecha: 31 de enero de 2026
Tema: Modelos predictivos y algoritmos de Machine Learning

■ Resumen de Respuestas

#	Respuesta	Confianza
1	**FALSO**	Alta
2	**VERDADERO**	Alta
3	**VERDADERO**	Alta
4	**FALSO**	Alta
5	**FALSO**	Alta

■ Análisis Detallado

Pregunta 1 (1 pt)

> **En los modelos de regresión de machine learning, la variable respuesta que se desea predecir es de tipo cualitativa.**

■ *Respuesta: FALSO*

Justificación:

En los modelos de **regresión**, la variable respuesta (variable dependiente o target) es de tipo **cuantitativa** (numérica continua), NO cualitativa.

- **Regresión** → Predice valores numéricos continuos (ej: precio de una casa, temperatura, tiempo de resolución de un caso judicial)
- **Clasificación** → Predice categorías/clases cualitativas (ej: spam/no spam, tipo de delito, diagnóstico médico)

Según Bonaccorso (2018), los algoritmos de regresión buscan encontrar una función que mapee las variables de entrada a un valor numérico continuo como salida.

Pregunta 2 (1 pt)

> **En aprendizaje supervisado, los árboles de regresión se caracterizan por la capacidad de modelar estructuras de regresión lineal.**

■ *Respuesta: VERDADERO*

Justificación:

Los árboles de regresión **SÍ tienen la capacidad** de modelar estructuras de regresión lineal, aunque no sean la herramienta más eficiente para ello.

La retroalimentación oficial del curso indica:

> *"Si las relaciones entre las variables de entrada y la variable respuesta se pueden aproximar por un modelo lineal, una regresión lineal funciona bien y supera a un árbol de regresión. Esto se debe a que un árbol de regresión no es capaz de modelar esta estructura [de manera óptima]."*

Punto clave: La pregunta se refiere a la **capacidad** de modelar, no a la **eficiencia** o superioridad. Los árboles de regresión pueden aproximar cualquier función (incluyendo lineales) mediante particiones sucesivas del espacio, aunque para relaciones puramente lineales, la regresión lineal clásica es más apropiada.

■■ **Nota de reflexión:** Mi error inicial fue confundir "capacidad de modelar" con "es la mejor opción para modelar". La capacidad existe; la eficiencia comparativa es otra cosa.

Pregunta 3 (1 pt)

> **Los modelos predictivos requieren de información de entrada para reconocer patrones.**

■ *Respuesta: VERDADERO*

Justificación:

Esto es un principio fundamental del machine learning. Los modelos predictivos necesitan:

1. **Datos de entrada (features/características)** para aprender patrones
2. **Datos de salida (labels/etiquetas)** en aprendizaje supervisado para aprender la relación entrada-salida

Sin información de entrada, el modelo no tiene nada que analizar ni patrones que reconocer. Es como intentar enseñarle a alguien a distinguir manzanas de naranjas sin mostrarle nunca una fruta.

En el contexto del foro de esta semana, las medidas de tendencia central son precisamente parte de esa información de entrada que ayuda al modelo a entender la distribución de los datos.

Pregunta 4 (1 pt)

> **Los modelos de aprendizaje automático denominados random forests se caracterizan por tener un bajo nivel de eficiencia al generar modelos predictivos.**

■ *Respuesta: FALSO*

Justificación:

Random Forest es conocido por su **ALTA eficiencia y precisión**, no baja. Sus características principales son:

- **Alta precisión** en la mayoría de problemas
- **Robustez ante overfitting** gracias al promedio de múltiples árboles
- **Manejo automático de valores faltantes** y variables categóricas
- **Reducción de varianza** mediante el ensamble de árboles decorrelacionados

Random Forest es uno de los algoritmos más populares precisamente porque logra buen rendimiento con poco ajuste de hiperparámetros. Según Breiman (2001), creador del algoritmo, RF combina la fuerza de múltiples árboles débiles para crear un predictor fuerte.

Pregunta 5 (1 pt)

> **En el clasificador de Bayes, los eventos a observar deben ser independientes y mutuamente excluyentes.**

■ *Respuesta: FALSO*

Justificación:

El clasificador **Naive Bayes** asume que las características (features) son **condicionalmente independientes** dado la clase, pero **NO** asume que sean **mutuamente excluyentes**.

- **Independencia condicional**: $P(A \cap B \mid C) = P(A \mid C) \times P(B \mid C)$
- **Mutuamente excluyentes**: Si ocurre A, entonces B no puede ocurrir (lo cual NO es un supuesto de Naive Bayes)

Por ejemplo, en clasificación de texto, el modelo asume que la presencia de la palabra "oferta" es independiente de la palabra "gratis" dado que el email es spam. Pero ambas palabras pueden aparecer juntas (no son mutuamente excluyentes).

El error en el enunciado está en confundir "independientes" con "mutuamente excluyentes", que son conceptos probabilísticos diferentes.

■ Referencias

- Bonaccorso, G. (2018). *Machine learning algorithms* (2nd ed., pp. 104-123). Packt Publishing.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Campesato, O. (2020). *Artificial intelligence, machine learning, and deep learning* (pp. 23-42). Mercury Learning & Information.

■ Relación con el Foro de la Semana 3

Estas preguntas complementan el tema del foro (medidas de tendencia central) porque:

1. Las medidas de tendencia central son **información de entrada** (Pregunta 3) que alimenta los modelos predictivos
2. Los modelos de **regresión** (Pregunta 1) predicen valores numéricos como los que analizamos con media/mediana
3. **Random Forests** (Pregunta 4) pueden usar estadísticas descriptivas como features para mejorar predicciones
4. Entender tipos de variables (cualitativas vs cuantitativas) es fundamental para elegir entre regresión y clasificación