



Campus Monterrey

Materia

Inteligencia artificial avanzada para la ciencia de datos II

Tarea

Capturas de pantalla del examen medio término

Estudiante

Jacobo Hirsch Rodriguez

A00829679@tec.mx

Profesor

Alfredo Esquivel Jaramillo

Este examen fue bloqueado en 31 de oct en 12:15.

enfoque que permite adaptar un modelo pre-trenado en una tarea específica utilizando menos datos y recursos. Seleccionó esta respuesta. Esta era la respuesta correcta.

MÁS RECIENTE	Intento	Hora	Puntaje
	Intento 1	57 minutos	76 de 100

Puntaje para este examen: **76** de 100
Entregado el 31 de oct en 12:12
Este intento tuvo una duración de 57 minutos.

Pregunta 18 / 8 pts

En el contexto del procesamiento de lenguaje natural, ¿qué describe mejor el concepto de "transfer learning"?

¡Correcto!

- ☒ El enfoque que permite adaptar un modelo pre-entrenado en una tarea específica utilizando menos datos y recursos.
- ☐ La práctica de entrenar un modelo desde cero utilizando un conjunto de datos muy grande.
- ☐ La técnica de dividir un corpus de texto en oraciones y palabras para su análisis.
- ☐ El proceso de crear modelos de lenguaje utilizando únicamente datos de un dominio específico.

Pregunta 28 / 8 pts

¿Cuál de las siguientes descripciones describe mejor el aprendizaje semisupervisado?

¡Correcto!

- ☒ Un método que combina datos etiquetados y no etiquetados, aprovechando ambos para mejorar el aprendizaje del modelo.
- ☐ Un enfoque que utiliza únicamente datos etiquetados para entrenar modelos de aprendizaje automático.
- ☐ Una técnica que se basa exclusivamente en la agrupación de datos sin utilizar etiquetas.
- ☐ Un enfoque que se centra en el aprendizaje a partir de interacciones con un entorno, como en juegos o simulaciones.

Pregunta 38 / 8 pts

Dado el siguiente texto en inglés: "Don't stop! Let's go to the park." ¿Cuál de los siguientes conjuntos de tokens es el resultado de aplicar cada uno de los siguientes tokenizadores?

A) WhiteSpaceTokenizer
B) WordPunctTokenizer
C) TreeBankWordTokenizer

WhiteSpaceTokenizer: ["Don't", "stop!", "Let's", "go", "to", "the", "park"]
WordPunctTokenizer: ["", "t", "stop", "!", "Let", "s", "go", "to", "the", "park", ""]
TreeBankWordTokenizer: ["Don", "", "t", "stop", "!", "Let", "s", "go", "to", "the", "park", ""]

Solo A es correcto.

¡Correcto!

- ☒ Todos los conjuntos son correctos.
- ☐ Solo A y B son correctos.
- ☐ Solo B y C son correctos.
- ☐ Solo A es correcto

Pregunta 40 / 8 pts

¿Cuál de las siguientes opciones de código en Python crea correctamente un vector de características utilizando el enfoque de Bag of Words para una lista de documentos?

Respuesta correcta

```
from sklearn.feature_extraction.text import CountVectorizer
documents = ["apple banana", "banana orange"]
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(documents)
print(X.toarray())
```

Respondido

```
documents = ["apple banana", "banana orange"]
word_cnt = {}
for doc in documents:
    for word in doc.split():
        word_cnt[word] = word_cnt.get(word, 0) + 1
print(word_cnt)
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
documents = ["apple banana", "banana orange"]
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(documents)
print(X.toarray())
```

Pregunta 5

8 / 8 pts

En el contexto de modelos de lenguaje basados en n-gramas, ¿cuál de las siguientes afirmaciones describe mejor la suposición markoviana y sus implicaciones en la construcción de un modelo de n-gramas?

¡Correcto!

- ☒ La suposición markoviana en un modelo de n-gramas sugiere que la probabilidad de cada n-ésima palabra depende de las n-1 palabras anteriores, lo que mejora la precisión del modelo al considerar más contexto.
- ☐ La suposición markoviana establece que la probabilidad de una palabra en una secuencia depende únicamente de la palabra anterior, lo que limita la capacidad del modelo para capturar dependencias a largo plazo en el texto.
- ☐ La suposición markoviana implica que la probabilidad de una palabra en una secuencia depende de todas las palabras anteriores en la oración, lo que permite que el modelo capture completamente el contexto semántico.
- ☐ En un modelo de n-gramas, la suposición markoviana permite que el modelo ignore por completo las palabras anteriores, enfocándose solo en la última palabra para predecir la siguiente.

Pregunta 6

0 / 8 pts

¿Cuál de las siguientes afirmaciones describe mejor el concepto de perplejidad en el contexto de modelos de lenguaje?

Respuesta correcta

Respondido

- ☒ La perplejidad es una medida de la cantidad de palabras únicas en un corpus, que indica la diversidad del vocabulario utilizado.
- ☐ La perplejidad mide la longitud promedio de las oraciones generadas por un modelo de lenguaje, siendo útil para evaluar la fluidez del texto.
- ☐ La perplejidad se utiliza para calcular la precisión de un modelo de lenguaje en tareas de clasificación, como el análisis de sentimientos.

Pregunta 8

14 / 14 pts

Considere el siguiente "toy example" basado en el siguiente corpus

<s> I am Sam </s>
<s> Sam I am </s>
<s> Sam I like </s>
<s> Sam I do like </s>
<s> do I like Sam </s>

Asuma que se usa un modelo de lenguaje de bigrama basado en esos datos de entrenamiento.

¿Cuál de las siguientes oraciones tiene una mayor probabilidad bajo este modelo?

- (1) <s> Sam I do I like </s>
- (2) <s> Sam I am </s>
- (3) <s> I do like Sam I am </s>

¡Correcto!

- ☒ 2
- ☐ 1
- ☐ 3

3.

Pregunta 9

0 / 8 pts

¿Cuál de las siguientes opciones describe mejor el concepto de "language modeling"?

Respuesta correcta

Respondido

- ☒ La técnica de predecir la probabilidad de una secuencia de palabras en un idioma.
- ☐ La tarea de asignar etiquetas a partes del habla en un texto.
- ☐ El proceso de identificar entidades nombradas dentro de un texto.
- ☐ La creación de resúmenes automáticos de textos largos.

Pregunta 10

8 / 8 pts

En el contexto del modelo skip-gram de Word2Vec, ¿cuál de las siguientes afirmaciones describe correctamente qué se utiliza como embeddings finales para las palabras después de entrenar el modelo?

¡Correcto!

☒

Los embeddings finales son los vectores de pesos de la capa oculta del modelo, que representan las representaciones distribucionales de las palabras y se obtienen a partir de la matriz de pesos que conecta la capa de entrada a la capa oculta.

☐

Los embeddings finales son los vectores de pesos de la capa de salida del modelo, que representan las probabilidades de las palabras de contexto para cada palabra central en el vocabulario.

☐

Los embeddings finales son generados a partir de la media de las representaciones de las palabras en el contexto, promediando los vectores de las palabras de entrada y salida durante el entrenamiento.

☐

Los embeddings finales son obtenidos mediante una técnica de reducción de dimensionalidad aplicada a la matriz de co-ocurrencia de palabras, que se usa para visualizar las relaciones semánticas entre las palabras.

Pregunta 11

8 / 8 pts

En PyTorch, ¿cuál es la forma correcta de mover un tensor desde la GPU al CPU y convertirlo en un array de NumPy?

(a) `tensor_cpu = tensor_gpu.cpu().numpy()`

(b) `tensor_cpu = tensor_gpu.to('cpu').numpy()`

(c) `tensor_cpu = tensor_gpu.detach().numpy()`

(d) `tensor_cpu = tensor_gpu.numpy().to('cpu')`

¡Correcto!

☒

A y B

☐

A

☐

B

☐

C

☐

D