

Tarea 0

Inteligencia Artificial



Lic. en Ciencias de la Computación

20 de enero de 2024 – 28 de enero de 2024

El objetivo de esta tarea es afilar tus habilidades en matemáticas y programación necesarias para este curso. Algunos de estos problemas van a aparecer en futuras tareas como subproblemas, así que asegúrate de entender cómo resolverlos. Si no sientes seguridad en responder algún problema o necesitas un repaso, revisa el material preeliminar de la guía de trabajo u otros recursos bibliográficos y digitales.

Instrucciones generales

Esta tarea tiene una parte escrita y una parte programada. Se utilizan los siguientes indicadores visuales:

-  Significa que debes escribir tus respuestas en `tarea0.pdf`.
-  Significa que debes escribir tus respuestas en `tarea0.py`.

Todas las respuestas escritas deben ser tipografiadas en \LaTeX . Se provee una plantilla en el material de ayuda `tarea0.zip`. Tus respuestas deben estar **en orden**, así como **clara y correctamente etiquetadas** para recibir el puntaje correspondiente.


Debes modificar el código de `tarea0.py` entre los marcadores `# INICIO` y `# FIN`. También puedes agregar funciones auxiliares fuera de estos delimitadores si gustas. No se permite cambiar otros archivos además de `tarea0.py`.

Tu código es evaluado con dos tipos de casos de prueba: las **básicas** y las **ocultas**, las cuales puedes ver en `califica.py`. Las pruebas básicas se proveen por completo y no utilizan entradas grandes o casos particularmente problemáticos. Las entradas de las pruebas ocultas también se proveen en `califica.py`, pero las salidas correctas no. Para correr las pruebas, puedes invocar el comando `python califica.py` en la terminal, el cuál te informa si pasaste las pruebas básicas y señala una alerta si tu código tarda demasiado o truena sin verificar la respuesta de las pruebas ocultas.

Se recomienda que leas y entiendas todos los casos de prueba y que agregues tus propios casos durante el desarrollo de la tarea.

Optimización y probabilidad

En este curso, trataremos muchos problemas de inteligencia artificial como problemas de optimización, es decir, buscaremos la mejor solución en un sentido matemático riguroso. Al mismo tiempo, deberemos lidiar con la incertidumbre del mundo, por lo que apelaremos a herramientas de probabilidad.


1.  Sean x_1, \dots, x_n números reales representando posiciones sobre una recta. Sean w_1, \dots, w_n números reales positivos representando la *importancia* de cada una de estas posiciones. Considera la función cuadrática,

$$f(\theta) = \sum_{i=1}^n w_i (\theta - x_i)^2$$

y que θ es un escalar. ¿Qué valor de θ minimiza $f(\theta)$? Muestra que el óptimo que encuentres es realmente un mínimo. ¿Qué cuestiones problemáticas pueden surgir si algunas de las w_i son negativas?

Puedes pensar en este problema como intentar encontrar el punto θ que no está muy lejos de las x_i .

Se espera que respondas con: Una expresión para el valor de θ que minimiza $f(\theta)$ así como una descripción de cómo la obtuviste; un cálculo o argumento corto que muestre que es un mínimo; una o dos oraciones describiendo algún problema que pueda surgir si alguna de las w_i son negativas.

2.  Considera las siguientes funciones,

$$f(\mathbf{x}) = \min_{s \in [-1, 1]} \sum_{i=1}^d s x_i$$

$$g(\mathbf{x}) = \sum_{i=1}^d \min_{s_i \in [-1, 1]} s_i x_i$$

donde $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ es un vector real y $[-1, 1]$ el intervalo cerrado entre -1 a 1 . ¿Cuál de las siguientes desigualdades es cierta para toda \mathbf{x} ? Demuéstralo.


$$f(\mathbf{x}) \leq g(\mathbf{x})$$

$$f(\mathbf{x}) = g(\mathbf{x})$$

$$f(\mathbf{x}) \geq g(\mathbf{x})$$


Puedes encontrar útil reescribir las funciones para que minimicen la misma cantidad, pero sobre conjuntos de tamaño distinto.

Se espera que respondas con: Una demostración breve, debes usar notación matemática, pero también puedes argumentar con palabras. Tu demostración puede ser algo informal, pero describe tus pasos y proceso de pensamiento claramente.

3.  Supongamos que lanzas repetidamente un dado justo de seis caras hasta que obtienes un resultado de 1 (y luego te detienes). Cada vez que lanzas un 3 ganas a puntos, y cada vez que lanzas un 6 pierdes b puntos. No ganas ni pierdes puntos si lanzas un 2, 4 o 5. ¿Cuál es la cantidad de puntos (como función de a y b) que esperamos tener cuando te detengas?

Puedes encontrar útil definir una recurrencia. Si defines V como la cantidad esperada de puntos al jugar el juego, ¿qué pasa si lanzas un 3? ganas a puntos y vuelves a lanzar el dado, ¿qué pasa en los otros casos? ¿puedes escribir esto como una recurrencia?

Se espera que respondas con: Una recurrencia que represente el problema y la expresión resultante de solucionar la recurrencia. Puedes resolver el problema de otras formas.

4.  Supongamos que la probabilidad de que una moneda caiga en águila es p (donde $0 < p < 1$), y que lanzas esta moneda cinco veces obteniendo (S, A, A, A, A) . Sabemos que la probabilidad de obtener esta secuencia es,

$$L(p) = (1 - p)p p p p = p^4(1 - p)$$

¿Qué valor de p maximiza $L(p)$? Muestra que este valor de p maximiza $L(p)$. ¿Cuál es una interpretación intuitiva de este valor de p ?

Considera tomar la derivada de $\log L(p)$. Puedes también tomar directamente la derivada de $L(p)$, pero puede resultarte menos simple hacerlo directamente. Puedes verificar que el valor de p que maximiza $\log L(p)$ también debe maximizar $L(p)$.

Se espera que respondas con: El valor de p que maximiza $L(p)$ y el cálculo realizado para encontrar la solución; una breve demostración que es un máximo. Un enunciado explicando la interpretación intuitiva de la solución.

5. ✍ Supongamos que A y B son dos eventos tales que $P(A \mid B) = P(B \mid A)$. También sabemos que $P(A \cup B) = \frac{1}{3}$ y que $P(A \cap B) > 0$. Muestra que $P(A) > \frac{1}{6}$.

Observa que A y B no necesariamente son mutuamente excluyentes. Considera cómo podemos relacionar $P(A \cup B)$ y $P(A \cap B)$.

Se espera que respondas con: Una breve demostración o derivación.

6. ✍ Considera un vector columna $\mathbf{w} \in \mathbb{R}^d$ y vectores columna constantes $\mathbf{a}_i, \mathbf{b}_j \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$ y un entero positivo n . Define la función con valor escalar,

$$f(\mathbf{w}) = \left(\sum_{i=1}^n \sum_{j=1}^n \left(\mathbf{a}_i^\top \mathbf{w} - \mathbf{b}_j^\top \mathbf{w} \right)^2 \right) + \frac{\lambda}{2} \|\mathbf{w}\|_2^2,$$

donde el vector es $\mathbf{w} = (w_1, \dots, w_d)^\top$ y $\|\mathbf{w}\|_2 = \sqrt{\sum_{k=1}^d w_k^2} = \sqrt{\mathbf{w}^\top \mathbf{w}}$ es conocida como la norma L_2 . Calcula el gradiente $\nabla f(\mathbf{w})$.

Recuerda que el gradiente es un vector d -dimensional de las derivadas parciales con respecto a cada w_i :


$$\nabla f(\mathbf{w}) = \left(\frac{\partial f(\mathbf{w})}{\partial w_1}, \dots, \frac{\partial f(\mathbf{w})}{\partial w_d} \right)^\top.$$

Si se te dificulta encontrar la respuesta, considera trabajar el problema utilizando escalares en lugar de vectores y derivadas en lugar de gradientes. No todo será similar, pero ambas respuestas deberán ser consistentes cuando $d = 1$. Procura no escribir sumatorias sobre las dimensiones ya que puede resultar muy tedioso.

Se espera que respondas con: Una expresión para el gradiente y el desarrollo realizado para derivarla. No expandas términos innecesariamente, intenta escribir la respuesta final de manera compacta.

Complejidad computacional

Al diseñar algoritmos, es útil poder realizar cálculos sobre servilleta para estimar qué tanto tiempo y espacio requiere un algoritmo. Afortunadamente, comenzarás a adquirir intuiciones para esto trabajando con distintos tipos de problemas.

1.  Supongamos que tienes una cuadrícula de puntos de $n \times n$, donde nos gustaría colocar 3 rectángulos alineados a los ejes (los lados del rectángulo son paralelos a los ejes). Cada esquina de cada rectángulo debe ser uno de los puntos en la cuadrícula, pero fuera de eso no hay restricciones sobre la ubicación o tamaño de los rectángulos. Por ejemplo, es posible que las cuatro esquinas de un rectángulo estén en el mismo punto (resultando en un rectángulo de tamaño 0), o que todos los 3 rectángulos estén encimados.

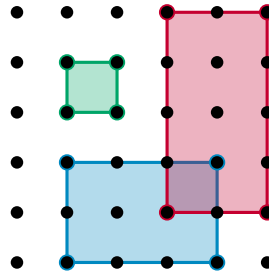



Figura 1: Ejemplo de un posible acomodo de rectángulos

¿De cuántas maneras se pueden colocar los 3 rectángulos sobre la cuadrícula? En general, solo nos importa la complejidad asintótica, entonces escribe tu respuesta de la forma $O(n^c)$ o de la forma $O(c^n)$ para algún entero c .

Es innecesario considerar si el orden importa en este problema, ya que se pregunta por complejidad asintótica. Eres libre de considerar cualquier orden en tu solución ya que no cambia la respuesta final.

Se espera que respondas con: Una cota superior asintótica en notación O grande de la cantidad de formas posibles de colocar 3 rectángulos y una simple explicación o razonamiento de tu respuesta.

2.  Supongamos que tienes una cuadrícula de puntos de $n \times 2n$. Comenzamos en el punto de la esquina superior izquierda (el punto en la posición $(1, 1)$), y nos gustaría llegar al punto de la esquina inferior derecha (el punto en la posición $(n, 2n)$) moviéndose exclusivamente o hacia abajo o hacia la derecha.

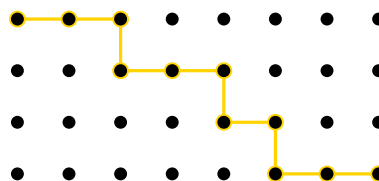


Figura 2: Ejemplo de un camino válido

Supongamos que se nos provee una función $c(i, j)$ que produce el costo asociado con la posición (i, j) , y supongamos que para cada posición toma tiempo constante calcular este costo. El costo puede ser negativo. Define el costo de un camino como la suma de $c(i, j)$ para todos los puntos (i, j) sobre el camino, incluyendo ambos extremos. Presenta un algoritmo para calcular el costo del camino de costo mínimo desde $(1, 1)$ hasta $(n, 2n)$ de la manera más eficiente posible (con la complejidad en tiempo más pequeña). ¿Cuál es el tiempo de ejecución?

Se espera que respondas con: Una descripción del algoritmo para calcular el costo del camino de costo mínimo lo más eficiente posible; el tiempo de ejecución en notación O grande y una breve explicación de cómo emerge del algoritmo.

Consideraciones éticas

Una de las metas de este curso es estudiar cómo atacar problemas reales con herramientas de inteligencia artificial. Sin embargo, los problemas reales tienen consecuencias reales. Además de habilidades técnicas, es importante desarrollar conciencia y perspicacia sobre cuestiones éticas asociadas con la inteligencia artificial.

En este problema, vas a analizar distintos escenarios utilizando los lineamientos éticos producidos por la conferencia NeurIPS. Los [lineamientos éticos de NeurIPS en su edición 2023](#) listan diecisiete preocupaciones no exhaustivas bajo los rubros de *Potential Negative Social Impacts* y *General Ethical Conduct* (las listas numeradas). Para cada escenario, vas a escribir un enunciado de un potencial impacto negativo. Para esto, primero debes determinar si el algoritmo, conjunto de datos, o técnica pudiera tener un impacto social negativo o violar la conducta ética general. Si este es el caso, lista una preocupación que viola y justifica por qué consideras que aplica al escenario. Si **no** consideras que el escenario tiene alguna preocupación ética, explica cómo llegaste a esta conclusión.

Cada uno de los escenarios fue extraído de un artículo real de investigación en inteligencia artificial. La ética de la investigación en esta área refleja fielmente las posibles consecuencias de desplegar inteligencia artificial.

Se espera que respondas con: Un párrafo por cada uno de los escenarios donde identifiques al menos una preocupación ética de los lineamientos éticos de NeurIPS y justifiques por qué consideras que aplica, o bien, una explicación de por qué consideras que no aplica ninguna preocupación ética.

Lineamientos de conducta ética general: Si se utilizan datos obtenidos de humanos, considera si esos datos podrían:

1. Contener cualquier información personalmente identificable o sensible. Por ejemplo, ¿el conjunto de datos utiliza características o información de etiquetas sobre nombres de individuos? ¿la gente dio su consentimiento para la recopilación de dichos datos? ¿podría el uso de los datos resultar degradante o vergonzoso para algunas personas?
2. Contener información que podría deducirse sobre personas que no han dado su consentimiento para compartir. Por ejemplo, un conjunto de datos para sistemas de recomendación podría revelar inadvertidamente información del usuario, como su nombre, dependiendo de las funcionalidades proporcionadas.
3. Codificar, contener o potencialmente exacerbar prejuicios contra personas de cierto género, raza, sexualidad o que tengan otras características protegidas. Por ejemplo, ¿representa el conjunto de datos la diversidad de la comunidad donde se pretende desplegar el sistema?
4. Contener experimentación con sujetos humanos y si ha sido revisada y aprobada por una junta de supervisión relevante. Por ejemplo, se espera que una junta ética revise los estudios que predicen características (e.g., estado de salud) a partir de datos humanos (e.g., contactos con personas infectadas por COVID-19).
5. Han sido desacreditados por los creadores. Por ejemplo, el conjunto de datos DukeMTMC-ReID fue retractado y no debe usarse en otros estudios.

También existen otras preocupaciones relacionadas con los datos que deben considerarse, incluyendo:

1. Consentimiento para utilizar o compartir los datos. Explicar si se ha pedido permiso al propietario de los datos para utilizarlos o compartirlos y cuál fue el resultado. Incluso si no se recibió el consentimiento, explicar por qué podría ser apropiado desde un punto de vista ético. Por ejemplo, si los datos se recopilaron de un foro público, ¿se pidió a sus usuarios que dieran su consentimiento para utilizar los datos que produjeron y, de no ser así, por qué?
2. Consideraciones específicas del dominio cuando se trabaja con grupos de alto riesgo. Por ejemplo, si la investigación implica trabajar con menores o adultos vulnerables, ¿se han implementado las salvaguardias pertinentes?
3. Filtrado de contenidos ofensivos. Por ejemplo, al recopilar un conjunto de datos, ¿cómo filtran los autores el contenido ofensivo como el lenguaje racista o las imágenes violentas?
4. Cumplimiento del RGPD y otras regulaciones relacionadas con los datos. Por ejemplo, si los autores recopilan datos obtenidos de humanos, ¿cuál es el mecanismo para garantizar el derecho de los individuos a ser olvidados (eliminados del conjunto de datos)?




Lineamientos de impactos sociales potencialmente negativos: Considera si los métodos y aplicaciones propuestos pueden:

1. Facilitar directamente el daño a los seres vivos. Por ejemplo: ¿podría integrarse en armas o sistemas de armas?
2. Plantear inquietudes sobre seguridad o protección. Por ejemplo: ¿existe el riesgo de que las aplicaciones puedan provocar accidentes graves o abrir vulnerabilidades de seguridad cuando se implementen en entornos del mundo real?
3. Plantear preocupaciones sobre derechos humanos. Por ejemplo: ¿podría usarse la tecnología para discriminar, excluir o impactar negativamente a las personas, incluidos impactos en la prestación de servicios vitales, como atención médica y educación, o limitar el acceso a oportunidades como el empleo?
4. Tener un efecto perjudicial sobre los medios de vida o la seguridad económica de las personas. Por ejemplo: ¿Tiene un efecto perjudicial sobre la autonomía, la dignidad o la privacidad de las personas en el trabajo o amenaza su seguridad económica (por ejemplo, mediante la automatización o la alteración de una industria)? ¿Podría utilizarse para aumentar la vigilancia de los trabajadores o imponer condiciones que presenten un riesgo para la salud y la seguridad de los empleados?
5. Desarrollar o ampliar formas dañinas de vigilancia. Por ejemplo: ¿podría usarse para recopilar o analizar datos de vigilancia masivos para predecir el estatus migratorio u otras categorías protegidas o usarse en cualquier tipo de elaboración de perfiles criminales?
6. Daña gravemente el medio ambiente. Por ejemplo: ¿la aplicación incentivaría daños ambientales significativos como la deforestación, la extracción de combustibles fósiles o la contaminación?
7. Engañar a las personas de maneras que causen daño. Por ejemplo: ¿podría utilizarse el enfoque para facilitar interacciones engañosas que causarían daños como robo, fraude o acoso? Considere los posibles daños que podrían surgir cuando la tecnología se utiliza según lo previsto, pero también aquellos que podrían surgir cuando la tecnología se utiliza indebidamente (intencionalmente o no).
8. Puede ser mal utilizado o modificado para producir resultados contradictorios. Por ejemplo, se ha demostrado que los enfoques basados en el aprendizaje automático para el descubrimiento de fármacos también son aplicables al diseño de armas bioquímicas.

Ejemplo de escenario: Un hospital de Estados Unidos recientemente implementó un nuevo programa de intervención que inscribe a pacientes en riesgo en programas para ayudarlos a abordar sus problemas médicos crónicos de manera proactiva antes de que los pacientes terminen en el hospital. El programa de intervención identifica automáticamente a los pacientes en riesgo al predecir las puntuaciones de riesgo de los pacientes, que se miden en términos de costos de aten-

ción médica. Sin embargo, observaron que para un nivel de puntuación de riesgo determinado, los pacientes negros están considerablemente más enfermos cuando se inscriben que los pacientes blancos, aunque la puntuación de riesgo de enfermedad asignada sea idéntica. Reasignaron manualmente las puntuaciones de riesgo de los pacientes en función de sus síntomas actuales y observaron que el porcentaje de pacientes negros que se inscribirían ha aumentado del 17% a más del 45%.

Ejemplo de solución: Es probable que este algoritmo haya codificado, contenga o potencialmente exacerbe los prejuicios contra personas de una determinada raza o etnia, ya que el algoritmo predice los costos de atención médica. Debido a que el acceso a la atención médica en los EE. UU. es desigual, los pacientes negros tienden a tener costos de atención médica más bajos que sus homólogos blancos. Por tanto, el algoritmo predecirá incorrectamente que tienen un riesgo menor.

1.  Una empresa de inversión desarrolla un modelo simple de aprendizaje automático para predecir si es probable que un individuo incumpla con un préstamo a partir de una variedad de factores, incluida la ubicación, la edad, la puntuación crediticia y los registros públicos. Después de examinar sus resultados, se encuentra que el modelo predice principalmente en función de la ubicación y que el modelo acepta principalmente préstamos de centros urbanos y niega préstamos a solicitantes rurales. Además, al observar el género y el origen étnico de los solicitantes, se encuentra que el modelo tiene una tasa de falsos positivos significativamente mayor para los solicitantes negros y masculinos que para otros grupos. En una predicción falsa positiva, un modelo clasifica erróneamente a alguien que no incumple como probable que incumpla.
2.  La estilometría es una forma de predecir la autoría de un texto anónimo o impugnado, mediante el análisis de los patrones de escritura en el texto anónimo y otros textos escritos por los autores potenciales. Recientemente, se han desarrollado algoritmos de aprendizaje automático de alta precisión para esta tarea. Si bien estos modelos se utilizan normalmente para analizar documentos históricos y literatura, podrían usarse para desanonimizar una amplia gama de textos, incluido el código.
3.  Un grupo de investigación analizó millones de rostros de celebridades de las imágenes de Google para desarrollar una tecnología de reconocimiento facial. Las celebridades no dieron permiso para que sus imágenes se utilizaran en el conjunto de datos y muchas de las imágenes tienen derechos de autor. Para fotografías con derechos de autor, el conjunto de datos proporciona enlaces URL a la imagen original junto con cuadros delimitadores para la cara.
4. Los investigadores han creado recientemente un modelo de aprendizaje automático que puede predecir especies de plantas automáticamente y directamente a partir de una sola fotografía. El modelo fue entrenado usando fotografías cargadas en una aplicación por







usuarios que dieron su consentimiento para usar sus fotografías con fines de investigación, y el modelo solo se usa dentro de la aplicación para ayudar a los usuarios a identificar plantas que podrían encontrar en la naturaleza.

Programación

En este problema, vas a implementar un bonche de funciones pequeñas. El propósito principal es que te familiarices con Python, pero adicionalmente, las funciones que vas a implementar serán de ayuda en futuras tareas.

No debes importar bibliotecas de terceros. Únicamente puedes utilizar módulos de la biblioteca estándar de Python.

Se espera que: Escribas código de Python implementando las funciones descritas en `tarea0.py`. Intenta escribir tu código tan limpio y simple como sea posible y asegúrate de escribir tus respuestas entre los delimitadores `# INICIO` y `# FIN`.

1.  Implementa `find_alphabetically_first_word` en `tarea0.py`.
2.  Implementa `euclidean_distance` en `tarea0.py`.
3.  Implementa `mutate_sentences` en `tarea0.py`.
4.  Implementa `sparse_vector_dot_product` en `tarea0.py`.
5.  Implementa `increment_sparse_vector` en `tarea0.py`.
6.  Implementa `find_nonsingleton_words` en `tarea0.py`.