# Reconocimiento de Patrones

# Redes Neuronales

# Metaheurísticas

## Algoritmos Heurísticos

Los algoritmos heurísticos son los más fáciles de utilizar, ya que se basan en el conocimiento de una heurística que guía el proceso de búsqueda. El conocimiento del problema usualmente ayuda a encontrar una heurística razonable que encontrará rápidamente una solución aceptable. Un algoritmo de este tipo sólo buscará dentro de un subespacio del área total a una solución buena (que no necesariamente es la mejor) que satisfaga las restricciones impuestas. La principal limitación es su incapacidad para escapar de óptimos locales encontrar soluciones parcialmente óptimas).

## Algoritmos Metaheurísticos

Una metaheurística es un proceso iterativo maestro que guía y modifica las operaciones de una heurística subordinada para producir eficientemente soluciones de alta calidad. Las metaheurísticas pueden manipular una única solución completa (o incompleta) o una colección de soluciones en cada iteración. La heurística subordinada puede ser un procedimiento de alto o bajo nivel, una búsqueda local, o un método constructivo. Entre los algoritmos metaheurísticos más conocidos están, el recocido simulado y los algoritmos genéticos.

Cuatro tipos:

1. Las metaheurísticas de relajación se refieren a procedimientos de resolución de problemas que utilizan relajaciones del modelo original (es decir, modificaciones del modelo que hacen al problema más fácil de resolver), cuya solución facilita la solución del problema original.
2. Las metaheurísticas constructivas se orientan a los procedimientos que tratan de obtener una solución a partir del análisis y selección paulatina de las componentes que la forman.
3. Las metaheurísticas de búsqueda guían los procedimientos que usan transformaciones o movimientos para recorrer el espacio de soluciones alternativas y explorar las estructuras de entornos asociadas.
4. Las metaheurísticas evolutivas están enfocadas a los procedimientos basados en conjuntos de soluciones que evolucionan sobre el espacio de soluciones.

Las metaheurísticas evolutivas establecen estrategias para conducir la evolución en el espacio de búsqueda de conjuntos de soluciones (usualmente llamados poblaciones) con la intención de acercarse a la solución óptima con sus elementos. El aspecto fundamental de las heurísticas evolutivas consiste en la interacción entre los miembros de la población frente a las búsquedas que se guían por la información de soluciones individuales.

Las diferentes metaheurísticas evolutivas se distinguen por la forma en que combinan la información proporcionada por los elementos de la población para hacerla evolucionar mediante la obtención de nuevas soluciones.

Los algoritmos genéticos y meméticos y los de estimación de distribuciones emplean fundamentalmente procedimientos aleatorios, mientras que las metaheurísticas de búsqueda dispersa o de reencadenamiento de caminos (Path-Relinking) emplean procedimientos sistemáticos.

## Metaheurísticas – Algoritmos Genéticos

En ocasiones la computación se basa en procesos observados de la naturaleza para resolver ciertos problemas: por ejemplo, las redes neuronales que replican los procesos de sinapsis entre las neuronas. En este caso, los algoritmos genéticos replican el modelo de selección natural propuesto por Darwin, y que resume la famosa frase ‘la supervivencia del más fuerte o adaptado’.

Este modelo básicamente dice que, dentro de una población, los individuos que sobreviven son aquellos que están más adaptados al medio, por lo tanto, las generaciones futuras de estos estarán mejor adaptadas ya que serán combinaciones de los mejores genes de sus antepasados. Además, esta teoría de la evolución introduce un concepto muy interesante que son las mutaciones. Una mutación es un pequeño cambio que se produce de manera aleatoria en ciertos individuos e introduce de esta manera versatilidad en las poblaciones. Habrá mutaciones que den lugar a cambios favorables y otros desfavorables.

Se utilizan para resolver problemas de Búsqueda y Optimización, ya que se basan en evolucionar poblaciones de soluciones hacia valores óptimos del problema.

**Estructura de un algoritmo genético**

Tener en cuenta:

* Individuo: los individuos de la población son las posibles soluciones al problema que se intenta resolver.
* Población: conjunto de individuos.
* Función fitness o de adaptación: función que evalúa a los individuos y les asigna una puntuación en función de que tan buenas sean las soluciones para el problema.
* Función de cruce: función que dados dos individuos genera dos descendientes a partir de la combinación de genes de sus padres, esta función depende del problema en cuestión.

Diagrama

Descripción generada automáticamente

* Fase inicial: se genera una población inicial de individuos (soluciones)
* Fase de evaluación: se evalúan los individuos de la población con la función fitness
* Fase de selección: se seleccionan los mejores individuos
* Fase de reproducción: se cruzan los individuos seleccionados mediante la función de cruce, dando lugar a una nueva generación que va a sustituir a la anterior
* Fase de mutación: se introducen mutaciones (pequeños cambios) en ciertos individuos de la nueva población de manera aleatoria o un entrecruzamiento cromosómico (también llamado crossover o recombinación)
* Se obtuvo una nueva generación, en general, con soluciones mejores que la anterior. Se vuelve al punto 2

Los algoritmos genéticos pueden finalizar cuando se alcanza un número de generaciones concreto o cuando cumplen una condición de parada.

**Ventajas**

Se desenvuelven bien en problemas con un paisaje adaptativo complejo: aquéllos en los que la función de aptitud es ruidosa, cambia con el tiempo, o tiene muchos óptimos locales, gracias a los cuatro componentes principales de los algoritmos genéticos, paralelismo, selección, mutación y cruzamiento, los que trabajan juntos para conseguir su buen desempeño.

Pueden explorar el espacio de soluciones en múltiples direcciones a la vez, por lo que, los algoritmos genéticos funcionan particularmente bien resolviendo problemas cuyo espacio de soluciones potenciales es realmente grande, demasiado vasto para hacer una búsqueda exhaustiva en un tiempo razonable

Los algoritmos genéticos realizan cambios aleatorios en sus soluciones candidatas y luego utilizan la función de aptitud para determinar si esos cambios producen una mejora. Como sus decisiones están basadas en la aleatoriedad, todos los caminos de búsqueda posibles están abiertos; en contraste a cualquier otra estrategia de resolución de problemas que dependa de un conocimiento previo

**Desventajas**

Si se elige mal una función de aptitud o se define de manera inexacta, puede que el algoritmo genético sea incapaz de encontrar una solución al problema, o puede acabar resolviendo el problema equivocado.

Pueden tardar mucho en converger, o no converger en absoluto, dependiendo en cierta medida de los parámetros que se utilicen tamaño de la población, el ritmo de mutación y cruzamiento, el tipo y fuerza de la selección.

El lenguaje utilizado para especificar soluciones candidatas debe ser robusto; es decir, debe ser capaz de tolerar cambios aleatorios que no produzcan constantemente errores fatales o resultados sin sentido. Una de las formas mas usadas es definir a los individuos como listas de números -binarios, enteros o reales- donde cada número representa algún aspecto de la solución candidata.

# Modelos Bayesianos

## Probabilidad a priori o incondicional

La probabilidad a priori o incondicional asociada a una proposición A es el grado de creencia que se le otorga en ausencia de cualquier otra información y se escribe como . Por ejemplo, dada la variable cuyo dominio (posibles valores que puede tomar) es , si la probabilidad a priori de que es , deberíamos escribir:

## Probabilidad condicional o a posteriori

Una vez que el agente obtiene alguna evidencia que afecta a la variable , las probabilidades a priori ya no son aplicables a . En su lugar usamos probabilidades a *posteriori o condicionales*. La probabilidad condicional del evento dada la ocurrencia del evento se escribe como . Por ejemplo, si la probabilidad de que dada la evidencia de que , deberíamos escribir o

Es la probabilidad de que ocurra un evento ***A*** dado que ocurre otro evento ***B*** (evidencia). No implica causa-efecto. Ejemplo:

Las probabilidades condicionales pueden definirse en t´erminos de probabilidades no condicionales. La ecuaci´on que la define es:

Donde es la probabilidad conjunta de y , es decir, la probabilidad de que y ocurran al mismo tiempo. De la ecuación anterior se obtiene:

Conocida como la *regla del producto*. Esta última ecuación es más intuitiva, ya que expresa que la probabilidad de que ocurran y es igual a la probabilidad de que ocurra dado , siempre y cuando ocurra , o sea, por la probabilidad de . La probabilidad condicional es una herramienta muy útil para representar información causal de la forma .

## Regla o teorema de Bayes

Me permite calcular una probabilidad condicional cuando tenemos la probabilidad condicional en sentido contrario.

Por ejemplo, si una alarma se dispara (por la razón que sea) con probabilidad , los robos ocurren con probabilidad y sabemos que la probabilidad de que nuestra alarma se dispare cuando hay un robo es , con la regla de Bayes podemos calcular la probabilidad de que esté ocurriendo un robo cuando suena la alarma, .

Un ejemplo para comprender la importancia del teorema de Bayes. Supongamos que Juan tiene un examen médico, que incluye una radiografía de tórax, como rutina de ingreso para su nuevo trabajo en un banco y que en la radiografía hay un hallazgo compatible con el cáncer de pulmón. A esto último lo representamos como . Al recibir el resultado, Juan piensa que tiene la enfermedad y se preocupa mucho, pero... ¿debería hacerlo? Sin conocer la exactitud del test, Juan realmente no puede saber qué tan probable es que tenga cáncer de pulmón. Cuando se entera que el test no es absolutamente concluyente, decide investigar y descubre que este tiene una tasa de falsos negativos de 0,4 y de falsos positivos 0,02. De los datos podemos deducir que:

Dadas estas probabilidades, Juan se siente un poco mejor. Sin embargo, nota que todavia no sabe cuál es la probabilidad de qué él tenga cáncer de pulmón. La probabilidad de que Juan tenga cáncer de pulmón es , y esta no es una de las probabilidades listadas recién.

Juan finalmente recuerda el teorema de Bayes y se da cuenta de que todavía necesita otra probabilidad para determinar la que le interesa. La probabilidad faltante es , que representa la probabilidad de que él tenga cáncer de pulmón antes de conocer el resultado de la radiografía. El otro dato útil que tiene Juan es que pertenece al grupo de personas que se realizó un examen preocupacional, no un examen motivado por síntomas. Entonces, cuando se entera de que solo 1 de cada 1000 nuevos empleados tiene cáncer de pulmón, asigna 0.001 a Ahora sí Juan aplica la regla de Bayes:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Entonces, ahora Juan sabe que su probabilidad de tener la enfermedad es cercana a y se relaja un poco mientras espera el resultado de otros estudios.

Supongamos que otra persona, Pedro, tiene el mismo diagnóstico de radiografía de tórax que tuvo Juan . Sin embargo, Pedro se hizo el estudio porque ha trabajado en minas durante 20 años, y sus empleadores se han preocupado porque notaron que cerca del 10% de sus trabajadores desarrollaron cáncer de pulmón después de trabajar varios años en las minas. ¿Cuál es la probabilidad de que Pedro tenga cáncer de pulmón? Basado en la información que tenemos sobre Pedro antes de que se realizara el test, le asignamos una probabilidad a priori Repitiendo los cálculos de la regla de Bayes para este valor, podemos concluir que para Pedro, no muy alentador.

## Distribución Conjunta Completa

La distribución conjunta completa es la distribución de probabilidad conjunta que considera el conjunto completo de variables. Es decir, que contiene la probabilidad de ocurrencia de cada una de las combinaciones posibles entre los valores que puede tomar cada variable. La tabla siguiente muestra la distribución conjunta completa para el caso sencillo del ejemplo de la radiografía de tórax de Juan.

Tabla

Descripción generada automáticamente

Notar que la última fila y columna contienen la probabilidad de ocurrencia de cada valor de y respectivamente. Estas probabilidades se llaman marginales y se calculan sumando los valores de la fila o columna donde se encuentran, es decir sumando todas las probabilidades conjuntas donde la variable toma el valor .

La siguiente tabla muestra la distribución conjunta completa para el ejemplo de Pedro.

Tabla

Descripción generada automáticamente

Podríamos unir los ejemplos, pero para eso deberíamos definir una nueva variable con dominio y conocer la probabilidad de que una persona trabaje en las minas. Si esto fuera asi, la tabla tendría todas las combinaciones de las tres variables.

La distribución conjunta completa es una tabla que contiene toda la información necesaria para responder cualquier pregunta sobre las variables o la combinación de ellas.

## Independencia

Dos eventos a y b son independientes si se cumple alguna de las siguientes condiciones:

Es decir, si y no son nulas, y son independientes cuando ). En ese caso, claramente la probabilidad de ocurrencia del evento no cambia si ocurre o no ocurre .

Solo si y son independientes, partiendo de y usando la regla del producto se puede ver que . Es importante tener en mente que la dependencia entre dos variables o eventos no implica que uno sea la causa del otro.

## Independencia Condicional

Dos eventos y son condicionalmente independientes dado , si y se cumple alguna de las siguientes afirmaciones:

1. y.

2. o .

Un ejemplo para comprender esta propiedad. Supongamos que en un pueblo existen dos vecinos que no tienen ningún tipo de interacción entre ellos. Cada vecino tiene cierta probabilidad de salir de su casa con paraguas cuando hay pronóstico de lluvias. Llamemos y a los eventos “salir con paraguas” para el vecino 1 y para el vecino 2 respectivamente. Naturalmente, y son altas en caso de pronóstico positivo. Cuando uno de los vecinos sale con paraguas es más probable que el otro lo haga también, es decir, y no son independientes. Esto no significa que se influyan mutuamente en el mundo real. El comportamiento se debe a que ambos eventos tienen la misma causa. Entonces, para un observador que no conoce el pronóstico, los eventos son dependientes. Si el observador conoce el pronóstico, los eventos se vuelven independientes dado el pronóstico. Es decir, si sabemos que va a llover, la probabilidad de ocurrencia del evento es condicionalemente independiente de la ocurrencia del evento y se escribe asi: , donde lluvia es el evento que representa al pronóstico positivo de lluvia.

La independencia condicional tiene particular importancia en la utilización de las redes bayesianas porque, como veremos más adelante, permiten representar toda la información necesaria mediante un conjunto reducido de probabilidades condicionales.

## Redes Bayesianas

Una red bayesiana es una estructura de datos que representa las dependencias entre variables. Muestra una descripción compacta de cualquier distribución de probabilidad conjunta completa. Es un grafo dirigido en el que cada nodo contiene información probabilística cuantitativa. Especificación completa:

1. Un conjunto de variables aleatorias forma los nodos de la red. Las variables pueden ser discretas o continuas.
2. Un conjunto de arcos dirigidos conecta pares de nodos. Si hay un arco de un nodo a un nodo , se dice que **X** es un padre de **Y**.
3. Cada nodo tiene una distribución de probabilidad condicionada que cuantifica el efecto de los padres del nodo.
4. El grafo no tiene ciclos dirigidos, entonces es un grafo acíclico dirigido, o GAD.

La topología de la red especifica las relaciones de independencia condicional que existen en el dominio. El significado intuitivo de un arco que sale de y apunta a es, habitualmente, que tiene una influencia directa sobre . Es generalmente sencillo para un experto del dominio decidir qué influencias directas existen en el área. Una vez que la topología de la red bayesiana está diseñada, necesitamos sólo especificar una distribución de probabilidad condicional para cada variable dados sus padres. La combinación de la topología y las distribuciones condicionales son suficientes para definir la distribución conjunta completa para todas las variables.

## Inferencia en Redes Bayesianas

La tarea básica de cualquier sistema de inferencia probabilista es calcular la distribución de probabilidad a posteriori para un conjunto de variables pregunta, dado algún evento observado (esto es, alguna asignación de valores para un conjunto de variables evidencia). Notación en este contexto:

es la variable pregunta

es el conjunto de variables evidencias **.** Y es un evento observado particular

denota las variables no evidencia (a veces llamadas variables ocultas).

Conjunto completo de variables: . Una pregunta típica pide la distribución de probabilidad a posteriori .

**Ejemplo**

Planteamos una pregunta sobre el ejemplo del robo del libro.

Diagrama

Descripción generada automáticamente

En la red del robo, podemos observar el evento en el que y . Podríamos entonces preguntarnos por la probabilidad de que haya ocurrido un robo: . Para hacer más cortas las expresiones, reemplacemos los nombres de las variables por:

La primera letra, por ejemplo para Robo. Como las variables de este ejemplo son binarias, vamos a usar la misma letra (en minúscula) para indicar el caso verdadero y la letra negada para el caso falso, por ejemplo para . Entonces, la pregunta se puede escribir como si queremos conocer tanto la probabilidad de que el robo haya ocurrido como la de que el robo no haya ocurrido (devuelve dos valores) y si sólo nos interesa la probabilidad de que el robo sea cierto.

**Usando las distribuciones conjuntas totales**

Si bien es el método más simple y directo para realizar inferencias sobre el dominio, técnicamente no es un método de inferencia sobre redes bayesianas porque requiere que las probabilidades estén expresadas como probabilidades conjuntas.

Almacenar probabilidades conjuntas no es eficiente. Se podrían calcular a partir de las probabilidades condicionales, pero esto no mejoraría la eficiencia. Además, aunque se partiera de las probabilidades conjuntas el método tiene complejidad algorítmica muy alta.

A pesar de estas desventajas, es importante comprender cómo y por qué funciona, ya que es la base para el resto de los métodos. La explicación está en la sección 13.4 del libro.

Con la regla de Bayes

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteTexto

Descripción generada automáticamente

No conocemos todavia Para calcularla suponemos que el robo existió, pero no sabemos nada sobre el terremoto, asi que debemos tener en cuenta las dos posibilidades (con y sin terremoto).

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

La probabilidad de que haya ocurrido un robo, dado que Mary y John llamaron para avisar, es cercana a 0.28.

## Algunas consideraciones

En las redes grandes que pueden aparecer en casos reales hay aspectos muy importantes a tener en cuenta:

* Llevar a cabo un proceso de deducción y análisis similar al del ejemplo anterior es prácticamente imposible.
* La representación de las probabilidades conjuntas totales tendría un tamaño inmanejable, por eso usamos las probabilidades condicionales.
* Los mecanismos de inferencia automático se deben enfocar en la eficiencia, hasta el punto donde, para casos de gran tamaño, el cálculo exacto es demasiado ambicioso.
* No es posible crear estas redes de forma manual. Tanto el conocimiento del dominio para definir la topología, como las probabilidades conocidas, no son suficientes. En estos casos, se deben utilizar métodos automáticos para aprender las redes.

**Inferencia Automática: ver libro**