



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital  
Universitat Politècnica de València

## **Boletín de ejercicios (III)**

Predicción Estructurada Estadística

*Autor:* Juan Antonio López Ramírez

Curso 2019-2020



## Ejercicio 1

---

En este ejercicio, se nos pide explicar brevemente las diferencias entre la clasificación y la predicción de salida estructurada, además de citar dos ejemplos de aplicación de cada paradigma.

La principal diferencia entre los dos es cómo manejar la salida. El problema de clasificación consiste en dar una muestra, tratando de conseguir la etiqueta que la representa. Las predicciones estructuradas son el conjunto de problemas cuyas variables de salida son mutuamente dependientes y, por tanto, el espacio de soluciones se estructura en una secuencia, gráfico, conjunto, etc. de posibles hipótesis.

Por un lado, el problema de determinar si un cierto correo electrónico es spam o no es un problema de clasificación. Esto depende del propio correo electrónico que se le pasa como entrada al sistema de clasificación.

Por otro lado, el análisis sintáctico es un ejemplo de problema de predicción estructurada. Este consiste en, dada una secuencia de entrada, construir un árbol cuyas hojas sean los elementos de la entrada y cuya estructura obedezca a cierta gramática. En el campo del Procesamiento del Lenguaje Natural (NLP por su nombre en inglés *Natural Language Processing*), este problema es bastante común.

## Ejercicio 2

---

En este ejercicio se nos pide justificar por qué la descomposición de naive Bayes de la ecuación 5 de las diapositivas es adecuada para el problema de reconocimiento de cariotipos.

El problema de reconocimiento de cariotipos simplificado establece lo siguiente: Dado un conjunto de 22 imágenes sin clasificar de cromosomas humanos, etiquetar cada imagen de un conjunto de 22 etiquetas, 1, 2, ... 22, de tal manera que cada etiqueta se asigna exactamente a una imagen. Supongamos que  $x$  sea la secuencia no ordenada de 22 cromosomas y  $h$  sea la secuencia de todas las 22 etiquetas.

Por lo tanto, la descomposición naive Bayes propuesta en la siguiente ecuación de las diapositivas:

$$P(x|h) = P(x_1, \dots, x_{22}|h_1, \dots, h_{22}) \approx \prod_{i=1}^{22} P(x_i|h_i) \quad (1)$$

es apropiada para el problema de reconocimiento de cariotipos porque la siguiente ecuación:

$$P(x|h) = P(x_1, \dots, x_{22}|h) = P(x_1|h)P(x_2|x_1, h) \dots P(x_{22}|x_1, x_2, \dots, x_{21}, h) \quad (2)$$

puede ser aproximada a esta:

$$\approx P(x_1|h_1)P(x_2|h_2) \dots P(x_{22}|h_{22}) \quad (3)$$

ya que existe independencia tanto en  $x$  como en  $h$ . La independencia en  $x$  existe porque la forma de un cromosoma no depende de la forma de otros cromosomas. Aunque, para facilitar la notación, el subíndice de la  $x$  esté en orden, dada la naturaleza de este problema no es estrictamente necesario que  $x_2$  venga después de  $x_1$ , o  $x_{15}$  antes de  $x_{16}$ .

Por eso, para reducir sustancialmente el coste de cálculo, es posible hacer la aproximación a la Ecuación 3.

Por otra parte, la independencia de la  $h$  existe porque la representación de un tipo determinado de cromosoma es completamente independiente de la representación de otros cromosomas. El razonamiento de esta independencia puede verse de manera similar a la independencia de los cromosomas. No existe un orden estricto de las etiquetas dadas en la historia pasada o futura.

Así que, gracias a ambas independencias, es posible abordar este problema con una descomposición naive Bayes.

### Ejercicio 3

En este ejercicio se nos pide explicar todos los pasos y asunciones necesarias para obtener la ecuación 9 de las diapositivas a partir de la ecuación 7.

Esta ecuación establece que, dada una muestra  $x$ , una historia  $h'$  y una cierta retroalimentación  $f$ , la hipótesis óptima es :

$$\hat{h} = \operatorname{argmax}_h P(h|x, h', f) \quad (4)$$

A partir de aquí es importante asumir un entorno de retroalimentación determinista. Esto nos va a permitir definir una función de decodificación que mapea cada señal de retroalimentación en su decodificación, es decir,  $d = d(f)$ . Esta suposición facilita mucho este problema ya que no es necesario tener un modelo de reconocimiento de la retroalimentación. Ahora es posible sustituir la realimentación  $f$  por su decodificación  $d$  como:

$$\hat{h} = \operatorname{argmax}_h P(h|x, h', d) = \operatorname{argmax}_h \frac{P(h, x, h', d)}{P(x, h', d)} \quad (5)$$

Como la probabilidad del denominador no depende de  $h$ , podemos obviarla al maximizar, por lo que la ecuación se nos queda tal que así:

$$\hat{h} = \operatorname{argmax}_h P(h, x, h', d) \quad (6)$$

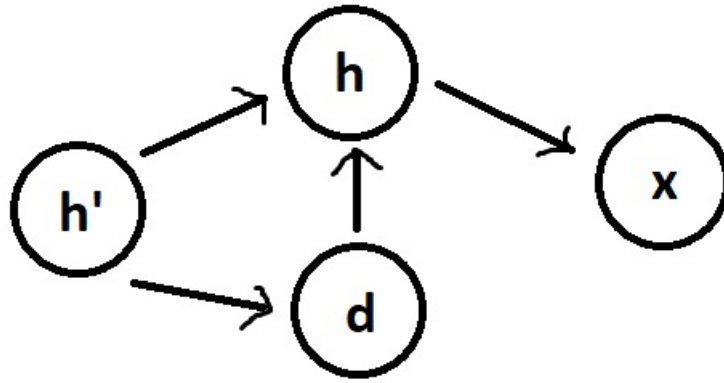
Una vez hecho esto, partimos de la ecuación que hemos obtenido para construir una red bayesiana que modele las relaciones entre los diferentes parámetros, tal y como se puede apreciar en la figura 1.

Partiendo de estas relaciones establecidas, la ecuación se nos queda de la siguiente forma:

$$\hat{h} = \operatorname{argmax}_h P(h, x, h', d) = \operatorname{argmax}_h P(h')P(d|h')P(h|h', d)P(x|h) \quad (7)$$

Finalmente, es posible simplificar la ecuación obtenida en el último paso, ya que  $h$  se obtiene tanto de  $h'$  como de  $d$  (de hecho  $d$  típicamente transmite información dirigida a modificar un elemento o una parte de  $h'$ ), por lo que podemos considerarlas independientes. Aplicando este último paso obtenemos:

$$\hat{h} = \operatorname{argmax}_h P(h')P(d|h')P(h|h', d)P(x|h) = \operatorname{argmax}_h P(x|h)P(h|h', d) \quad (8)$$



**Figura 1:** Red bayesiana que modela las relaciones entre nuestras variables.

Como podemos observar, el resultado de aplicar todos los pasos que se han explicado en este ejercicio es el mismo que la ecuación 9 de las transparencias.

## Ejercicio 6

Ahora, partiendo de la ecuación 7, hay que explicar como se consigue la ecuación 19 de las transparencias, de forma parecida a como lo hemos hecho en el ejercicio anterior. La diferencia es que aquí no asumiremos que la retroalimentación va a ser determinista. Partiendo de:

$$\hat{h} = \operatorname{argmax}_h P(h|x, h', f) = \operatorname{argmax}_h \frac{P(h, x, h', f)}{P(x, h', f)} \quad (9)$$

Es posible sacar el denominador, ya que es independiente de la variable que estamos tratando de maximizar. Ahora, necesitamos tener en cuenta toda la decodificación posible para una  $f$  proporcionada. Podemos añadir la variable de decodificación  $d$ , por lo que la ecuación se nos queda como:

$$\hat{h} = \operatorname{argmax}_h \sum_d P(h, x, h', f, d) \quad (10)$$

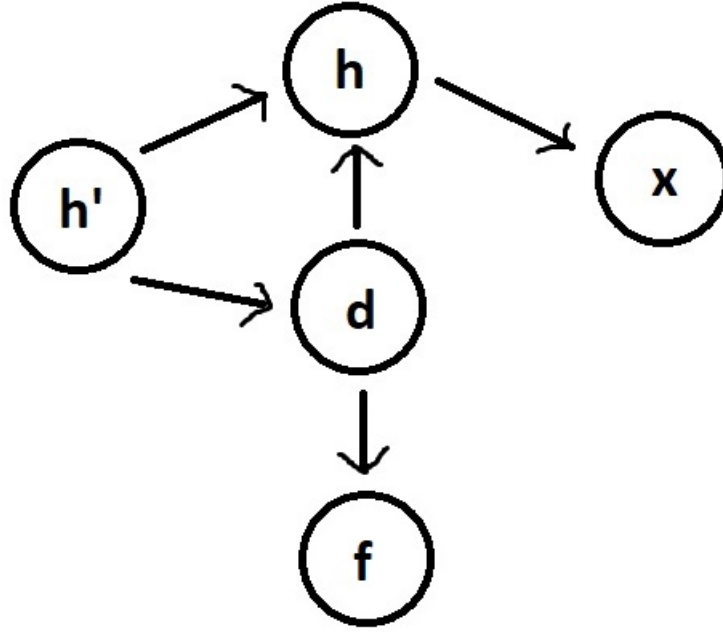
Ahora, de manera similar a como hemos hecho en el ejercicio 3, vamos a estructurar nuestra red bayesiana que establezca las relaciones entre nuestras variables. El resultado de esto se puede apreciar en la figura 2.

Teniendo en cuenta estas relaciones, obtenemos:

$$\hat{h} = \operatorname{argmax}_h \sum_d P(h')P(d|h')P(f|d)P(h|h', d)P(x|h) \quad (11)$$

Ahora, simplificando y sacando tanto las probabilidades independientes como el factor común, podemos obtener:

$$\hat{h} = \operatorname{argmax}_h P(x|h) \sum_d P(d|h')P(f|d)P(h|h', d) \quad (12)$$



**Figura 2:** Red bayesiana que modela las relaciones entre nuestras variables.

Finalmente, tratamos de encontrar la hipótesis óptima para una entrada  $x$  y algo de retroalimentación  $f$ , junto con la decodificación óptima para la retroalimentación dada  $f$ , obteniendo de esta forma:

$$(\hat{h}, \hat{d}) = \underset{h, d}{\operatorname{argmax}} P(d|h')P(f|d)P(x|h)P(h|h', d) \quad (13)$$

Como podemos observar, el resultado de aplicar todos los pasos que se han explicado en este ejercicio es el mismo que la ecuación 19 de las transparencias.

## Ejercicio 7

En este apartado vamos a explicar, brevemente, bajo qué condiciones la solución dada por las ecuaciones 22 y 23 de las transparencias puede ser óptima, además de si se mantienen las mismas condiciones para la optimización de la solución dada por las ecuaciones 20 y 21. Para ello, utilizaremos el ejemplo del cariotipo para ilustrar nuestras respuestas.

Las ecuaciones 22 y 23 de las diapositivas pueden ser óptimas en el caso de que  $n$  sea el número de clases del problema. Por ejemplo, en el problema del cariotipo, esas ecuaciones serían óptimas si  $n = 22$ , que es el número de cromosomas. Esto se debe a que en dichas ecuaciones se tiene en cuenta un conjunto de  $n$  decodificaciones. Si se considera el mismo número de decodificaciones que el número de clases, es posible tener la retroalimentación completa para corregir todos los errores. Por lo tanto, es posible calcular la hipótesis óptima.

Por otro lado, las ecuaciones 20 y 21 nunca serán óptimas con las mismas condiciones, ya que, en ellas, se obtiene primero la decodificación "óptima" para la retroalimentación  $y$ , con esa decodificación fija, se obtiene la hipótesis "óptima". El problema de este método es que puede existir una combinación "no óptima" de decodificaciones que permita obtener una mejor hipótesis como subproducto de ambas variables.

## Ejercicio 8

---

Aquí se nos pide explicar brevemente los conceptos y las principales diferencias entre los protocolos de interacción activos y pasivos. La principal diferencia entre ambos es quién decide qué elemento de la hipótesis debe ser supervisado.

En el protocolo de interacción pasiva, el operador humano tiene la iniciativa cuando se supervisa un elemento de una hipótesis. Este protocolo se puede dividir en dos principales tipos en cuanto al orden de supervisión, el de **Left-to-Right** y el **Desultory**. Los protocolos pasivos de Left-to-Right son aquellos en los que los elementos de la hipótesis son supervisados en un orden fijo. Concretamente, de izquierda a derecha, como su propio nombre indica. Este tipo de supervisión permite que el sistema asuma que la parte de más a la izquierda de la hipótesis va a ser correcta y que modifique la parte situada más a la derecha con cada supervisión del usuario. Por otro lado, la supervisión Desultory es aquella en la que el usuario puede modificar el elemento deseado sin un orden estricto. Esta supervisión permite que el usuario elija modificar el error más importante en cada paso de la interacción.

El protocolo interactivo activo propone un sistema capaz de tomar la iniciativa y ser el que le diga al usuario que supervise un elemento de la hipótesis. En cada paso de la interacción, el sistema debe calcular alguna medida de confianza para cada elemento. Un enfoque podría ser que el elemento de confianza más bajo se propone para la supervisión. De esta forma, el usuario valida si el elemento es correcto o necesita una corrección. Con esta modificación, el sistema calcula las próximas predicciones en base a la nueva retroalimentación y el historial previo.