

José Javier Calvo Moratilla
2021/2022

Predicción Estructurada Estadística

Entrega nº3

Ejercicios teóricos:

Ejercicios realizados: 1, 2, 3, 6, 7, 8

Question1: Briefly explain the differences between Classification and Structured Output Prediction. Cite two application examples for each paradigm.

El paradigma de clasificación consiste en indicar la clase de pertenencia de una muestra, normalmente identificada por la variable x , sobre un conjunto de clases C , dónde a la salida se obtiene la probabilidad de pertenencia a cada una de las clases previamente definidas y que posteriormente, mediante el uso de la operación ($\arg \max$) se puede obtener la clase más probable como hipótesis para una muestra x .

La variable x puede contener el conjunto de píxeles de una imagen en una tarea de clasificación de animales, o un conjunto de información relativa a un tipo de flor, como el tamaño de los pétalos, el color, para así clasificar tipos de flores.

El paradigma de la predicción estructurada estadística, en comparación con el de clasificación, tiene que decidir por un número de clases elevado, por ello el desarrollo de dichas hipótesis se complica, las exigencias computacionales son más significativas, por ello se utilizan los algoritmos vistos en la asignatura, como Viterbi o Inside-Outside.

Otro de los aspectos de la predicción estructurada estadística es que las variables de salida son dependientes entre sí.

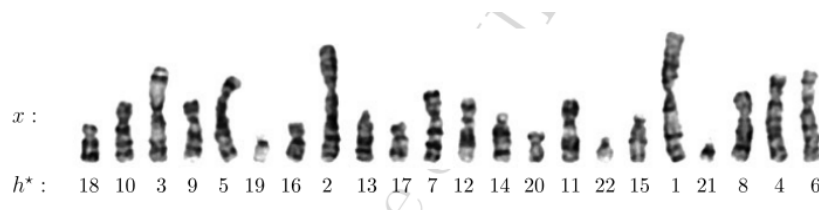
Una de las ventajas de dicho paradigma en comparación con el de clasificación es que pueden conseguir realizar hipótesis de buena calidad con un conjunto de datos reducido de ejemplo.

Es el paradigma utilizado en el laboratorio de la asignatura, en el que la muestra de

entrada x está formada por una secuencia de símbolos o palabras estructuradas y se puede aplicar en el procesamiento del lenguaje natural, por ejemplo en tareas de POS tagging, o en el campo de la genética para intentar encontrar algún patrón relevante de un conjunto cadenas genéticas.

Question2: Justify why the naive Bayes decomposition of Eq.(5) is adequate for karyotype recognition problems.

Partimos de imágenes de cromosomas individuales que pueden etiquetarse en 22 clases (h) diferentes.



Las variable x son la representación en píxeles de las imágenes de los cromosomas y la h su etiqueta correspondiente.

Para la que la descomposición de naive Bayes sea adecuada para el reconocimiento de cariotipos, debe de cumplirse, que siempre y cuando la presencia o ausencia de una característica determinada no está relacionada con la presencia o ausencia de cualquier otra característica.

Por tanto partimos de la siguiente ecuación inicial de naive Bayes:

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

Dicha ecuación es apropiada para el problema de reconocimiento de cariotipos ya que se puede expresar mediante la siguiente ecuación:

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

Y por ello se puede aproximar a:

$$p(x | h) \approx p(x_1 | h_1) \cdot P(x_2 | h_2) \dots P(x_{22} | h_{22})$$

Con la ecuación anterior se concluye que la estimación de los cariotipos no es dependiente de hipótesis anteriores, por ello se reduce el coste computacional del algoritmo utilizando la asunción de naive Bayes.

Question3: Briefly explain all the steps and assumptions needed to derive Eq.(9) from Eq.(7).

Partiendo de la ecuación nº 7 se deriva para obtener la ecuación nº9

$$\hat{h} = \operatorname{argmax}_{h \in H} P(h | x, h', f)$$

En primer lugar la ecuación inicial se reescribe utilizando el teorema de bayes:

$$P(A | B, C) = \frac{P(A, B, C)}{P(B, C)}$$

$$\hat{h} = \operatorname{argmax}_{h \in H} \frac{P(h, x, h', f)}{P(x, h', f)}$$

La variable h' identifica al histórico, la variable f a un feedback. El feedback puede sustituirse por la función d, que consiste en decodificar del feedback.

Dado que la variable h no se encuentra en $P(x, h', f)$, para la operación de argmax es una constante, por ello se puede obviar de la ecuación:

$$\hat{h} = \operatorname{argmax}_{h \in H} P(h, x, h', f)$$

Para el siguiente paso se utiliza la regla de la cadena y la función de decodificación $d = d(f)$, obteniendo la siguiente ecuación:

$$\hat{h} = \operatorname{argmax}_{h \in H} P(d) \cdot P(h' | d) \cdot P(h | h', d) \cdot P(x | h', d, h)$$

Dado que hay términos que no dependen de h como $P(d)$, $P(h' | d)$ para calcular la operación argmax, se eliminan de la ecuación:

$$\hat{h} = \operatorname{argmax}_{h \in H} P(h | h', d) \cdot P(x | h', d, h)$$

En último lugar, $P(x | h', d, h)$ Se considera independiente de h' y d dado h, ya que la nueva hipótesis contiene las modificaciones introducidas por la señal de feedback sobre la hipótesis anterior, obteniendo la ecuación nº9.

$$\hat{h} = \underset{h \in H}{\operatorname{argmax}} P(h | h', d) \cdot P(x | h)$$

Question6: Briefly explain all the steps and assumptions needed to derive Eq.(19) from Eq.(7).

Se parte de nuevo de la ecuación nº7 para obtener la ecuación nº19.

$$\hat{h} = \underset{h \in H}{\operatorname{argmax}} P(h | x, h', f)$$

De nuevo la variable x representa a una muestra, la variable h' el histórico, y la f un feedback.

Se utiliza la regla de bayes para transformar la ecuación obteniendo:

$$\hat{h} = \underset{h \in H}{\operatorname{argmax}} \frac{P(h, x, h', f)}{P(x, h', f)}$$

En dicho caso el denominador $P(x, h', f)$ no depende de h por ello se omite. El siguiente paso es construir un sumatorio para cubrir todas las combinaciones posibles de decodificaciones de feedback existentes, obteniendo la siguiente ecuación:

$$\hat{h} = \underset{h \in H}{\operatorname{argmax}} \sum_d P(h, x, h', f, d) = \underset{h \in H}{\operatorname{argmax}} \sum_d P(h') \cdot P(d|h') \cdot p(f|d) \cdot p(h|h', d) \cdot p(x|h)$$

Simplificamos la ecuación tomando las probabilidades independientes y sacando el factor común, obteniendo la siguiente ecuación:

$$\hat{h} = \underset{h \in H}{\operatorname{argmax}} P(x|h) \cdot \sum_d P(d|h') \cdot P(f|d) \cdot P(h|h', d)$$

Como último paso se aproxima la suma con la moda y se ignoran las probabilidades que no dependen de las variables h y d, obteniendo la ecuación:

$$(\hat{h}, d) \approx \underset{h, d}{\operatorname{argmax}} P(f|d) \cdot P(d|h') \cdot P(f|d) \cdot P(x|h) \cdot P(h|h', d)$$

Question7: Briefly explain under which conditions the solution given by Eq.(22-23) may be optimal. Do the same conditions hold for the optimality of the solution given by Eq.(20-21)? Why? Use the karyotyping example to illustrate your (otherwise general) responses.

Se puede considerar una solución óptima si se cumple que la variable n es igual al tamaño del problema. Si se observa en problema anterior de los cariotipos, si la variable n es igual a 22 (Tamaño de problema).

Al tratar un conjunto de n decodificaciones ecuanímes con el tamaño del problema en cuestión, se puede tener una retroalimentación completa para la corrección total de los errores producidos, pudiendo ser así calculada una hipótesis que maximiza la optimización.

Las ecuaciones 20,21 no podrán ser óptimas en base a estas condiciones, ya que éstas calculan primero la decodificación óptima para la retroalimentación y obtienen así consigo la hipótesis óptima.

Question8: Briefly explain the concepts and main differences between Active and Passive interaction protocols.

En primer lugar se identifican los protocolos de interacción activos, que consiste en otorgar al sistema el poder de toma de decisiones, para decidir de manera autónoma qué hipótesis deben de ser revisadas por el ser humano.

La calidad de los resultados depende del sistema y la métrica que utiliza para medir la confianza que se tiene en cada una de las hipótesis realizadas, y así decidir qué hipótesis necesitan ser o no revisadas por parte del ser humano.

Uno de los puntos fuertes de dicho protocolo es que el ser humano sólo analiza un subconjunto reducido de las hipótesis totales generadas, no tiene que revisar la totalidad de ellas, que puede ser un problema si la cantidad de hipótesis es muy elevada.

Por otra parte los protocolos de interacción pasivos consisten en otorgar a una persona experta el poder en la toma de decisiones, dónde decide qué hipótesis deben o no supervisarse.

Que el ser humano tome la decisión no es del todo negativo, ya que si una persona experta conoce muy bien una tarea concreta, sus revisiones pueden ser

determinantes para obtener un buen rendimiento en el sistema.

La revisión de hipótesis pasiva se puede clasificar de dos maneras diferenciadas:

Primero se observan las Left-to-right, cuyas revisiones se realizan en un orden predefinido y en segundo lugar las Desultory, cuyas revisiones se realizan de manera aleatoria.