

Predicción estructurada estadística: 3º boletín

Marcos Esteve Casdemunt

Febrero 2020

Contents

1	Ejercicio 1	2
2	Ejercicio 2	2
3	Ejercicio 3	3
4	Ejercicio 6	3
5	Ejercicio 8	4

1 Ejercicio 1

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

Clasificación: En este caso se trata de un problema donde hay que clasificar una determinada muestra en un conjunto finito y normalmente pequeño de clases. El problema se reducirá por tanto a realizar una exploración exhaustiva con el fin de obtener aquella clase que maximiza la probabilidad de que la muestra pertenezca a dicha clase. Podemos representarlo formalmente como:

$$\arg \max_{c \in C} P(c|x)$$

Donde C es el conjunto reducido de clases disponible y x es la muestra que quiere clasificarse.

Algunos ejemplos de una tarea de clasificación podrían ser:

- Detección de SPAM o no SPAM en correos electrónicos
- Detección de dígitos manuscritos aislados (MNIST)

Predicción estructurada estadística: La principal diferencia que existe con un problema de clasificación es que el espacio de hipótesis es potencialmente infinito y además el resultado de la muestra a analizar esta estructurada en una secuencia de símbolos, un grafo etc. Además destacar que al tratarse de problemas más complejos es necesario el uso de técnicas búsqueda con un coste computacional más elevado que en los problemas de clasificación. Tal y como se ha comentado podemos formalizarlo con la siguiente ecuación

$$\arg \max_{c \in C} P(c|x)$$

Donde a diferencia del problema de clasificación, C es un conjunto generalmente muy grande de clases.

Algunas aplicaciones de la predicción estructurada estadística son:

- Resumen automático, donde dado un texto de entrada se quiere generar un resumen del mismo.
- Dado un audio generar los subtítulos del mismo.

2 Ejercicio 2

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

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

La asunción naive Bayes parece ser adecuada a la hora de estimar los cariotipos debido a que permite evitar estimar una gran cantidad de parámetros y de esta

forma reducir el coste computacional del algoritmo. Para ello se considera que la hipótesis de cada cromosoma solo depende de si mismo y no del resto de cromosomas de la secuencia, por lo que la estimación se realizará utilizando modelos más simples.

3 Ejercicio 3

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

$$\hat{h} = \arg \max_{h \in H} P(h|x, h', f)$$

Podemos reescribirla utilizando la regla conjunta

$$\hat{h} = \arg \max_{h \in H} \frac{P(h, x, h', f)}{P(x, h', f)}$$

Como podemos observar el dividendo no depende del parámetro h que estamos maximizando por lo que resultará constante. Por tanto nos quedará la ecuación como:

$$\hat{h} = \arg \max_{h \in H} P(h, x, h', f)$$

Utilizando la regla de la cadena y utilizando la función de decodificación $d = d(f)$, la cual nos permite interpretar la función de feedback, podemos obtener la siguiente ecuación

$$\hat{h} = \arg \max_{h \in H} P(d)P(h'|d)P(h|h', d)P(x|h', d, h)$$

De nuevo $P(d)$ y $P(h'|d)$ no dependen del parámetro que estamos maximizando (h) y por tanto serán constantes y podremos sacarlos fuera de la ecuación.

$$\hat{h} = \arg \max_{h \in H} P(h|h', d)P(x|h', d, h)$$

Por último, de acuerdo con lo expuesto en las diapositiva, podemos observar que $P(x|h', d, h)$ puede considerarse 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. La ecuación quedara por tanto como:

$$\hat{h} = \arg \max_{h \in H} P(h|h', d)P(x|h)$$

4 Ejercicio 6

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

Partiendo de la ecuación

$$\hat{h} = \arg \max_{h \in H} P(h|x, h', f)$$

Podemos derivar la siguiente ecuación añadiendo la decodificación y marginalizando para todo d

$$\hat{h} = \arg \max_{h \in H} \sum_d P(h, d | x, h', f)$$

Utilizando la regla conjunta podemos derivar la ecuación

$$\hat{h} = \arg \max_{h \in H} \sum_d \frac{P(h, d, x, h', f)}{P(x, h', f)}$$

Como podemos observar el denominador no depende del parámetro que estamos maximizando y por tanto podemos sacarlo ya que será constante

$$\hat{h} = \arg \max_{h \in H} \sum_d P(h, d, x, h', f)$$

Se puede aproximar el sumatorio con la moda por lo que la ecuación quedará como

$$\hat{h} \approx \arg \max_{h \in H} \max_d P(h, d, x, h', f)$$

Utilizando la regla de la cadena podemos obtener la siguiente ecuación

$$\hat{h} \approx \arg \max_{h \in H} \max_d P(h') P(d | h') P(f | h', d) P(h | h', d, f) P(x | h', d, f, h)$$

Como podemos observar $P(h')$ no depende de los parámetros que estamos maximizando, por lo que será constante y podremos sacarlo fuera.

En $P(f | h', d)$ la decodificación depende del usuario al ver la historia pasada por lo que tendrá incluida información de dicha historia.

En $P(h | h', d, f)$ la decodificación tiene incluida la información del feedback, pudiendo esta señal de feedback eliminarse.

Por último en $P(x | h', d, f, h)$ la nueva hipótesis tiene la información de la historia pasada, la decodificación de la señal de feedback y la señal de feedback

Con todo esto la ecuación queda como:

$$\hat{h} \approx \arg \max_{h \in H} \max_d P(d | h') P(f | d) P(h | h', d) P(x | h)$$

5 Ejercicio 8

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

En el protocolo de interacción **pasivo** el usuario experto decide que muestras necesitan de supervisión. Por lo que el experto deberá de revisar todas las muestras y esto supondrá que el sistema sea perfecto desde su punto de vista. Las hipótesis del sistema se suelen revisar de dos formas distintas:

- **Left-to-right (Izquierda a derecha)** donde se revisan las hipótesis del sistema en un orden determinado.
- **Desultory (Aleatorio)** en caso de revisarse las hipótesis del sistema de forma aleatoria

Por otra parte el protocolo de interacción **activo** consiste en ceder la responsabilidad de decisión de que hipótesis deben ser revisadas al sistema. Una de las principales ventajas que supone este protocolo de interacción radica en que el operador humano solo analizará un subconjunto del total de las hipótesis generadas por el sistema. Destacar que la calidad de los resultados dependerá de la calidad del sistema para decidir que muestras necesitan supervisión. Por tanto, el sistema deberá de computar una medida de confianza para cada hipótesis y pedirle al usuario que evalúe aquellas muestras para las que el sistema tiene una menor confianza.