PRÁCTICO II

ESTIMACIÓN PARAMÉTRICA BAYESIANA Y POR MÁXIMA VEROSIMILITUD

Ejercicios Teóricos

Ejercicio 1. Tenga x una densidad exponencial

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases}$$

a) Grafique $p(x|\theta)$ versus x para $\theta = 1$. Grafique $p(x|\theta)$ versus θ, $(0 \leq \theta \leq 5)$, para $x = 2$.

b) Suponga que n ejemplos $x_1, \cdots, x_n$ se generan independientemente de acuerdo a $p(x|\theta)$. Muestre que el estimador de máxima verosimilitud para θ está dado por

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$$

c) En su gráfico generado con $\theta = 1$ en la parte a), marque el estimador de máxima verosimilitud $\hat{\theta}$ para n grandes.

Ejercicio 2. Tenga x una densidad uniforme

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & \text{si } 0 \leq x \leq \theta \\ 0 & \text{en caso contrario} \end{cases}$$

a) Suponga que n ejemplos $D = \{x_1, \ldots, x_n\}$ son generados independientemente de acuerdo a $p(x|\theta)$. Muestre que el estimador de máxima verosimilitud para θ es $max[D]$, en el caso que el valor del máximo elemento esté en D.

b) Suponga que son gerados $n = 5$ puntos de distribución y el máximo valor pasa a ser $max_k(x_k) = 0.6$. Grafique la verosimilitud $p(D|\theta)$ en el rango $0 \leq \theta \leq 1$. Explique con palabras por qué no es necesario conocer los otros 4 puntos.

Ejercicio 3 Demuestre que, si nuestro modelo es pobre, el clasificador de máxima verosimilitud al que arribamos no es el mejor, incluso entre nuestro conjunto de modelos. Supóngase que tenemos dos categorías (es decir $P(w_1) = P(w_2) = 0.5$). Además conocemos que $p(x|w_1) \sim N(0,1)$ pero asuma que $p(x|w_2) \sim N(\mu,1)$. Esto es, el parámetro θ que buscamos por técnicas de máxima verosimilitud es la media de la segunda distribución. Sin embargo, suponga que la distribunión subyacente es $p(x|w_2) \sim N(1,10^6)$.

a) Cuál es el valor de máxima verosimilitud estimado $\hat{\mu}$ en nuestro modelo pobre, dado una gran cantidad de datos?

b) Cuál es la frontera de desición a la que arribamos con el estimador de máxima verosimilitud de nuestro modelo pobre?

c) Ignore, por el momento, la aproximación de máxima verosimilitud y use el método del capítulo 2 para deducir los límites de decisión óptimos del teorema de Bayes dada la verdadera distribución subyacente $p(x|w_1) \sim N(0,1)$ y $p(w|w_2) \sim N(1,10^6)$. Tenga el cuidado de incluir todas las partes de las fronteras de decisión.

d) Ahora considere nuevamente los clasificadores basados en el modelo pobre de $p(x|w_2) \sim N(\mu, 1)$. Usando el resultado inmediatamente superior, encuentre un nuevo valor de μ que dé un error menor que el clasificador de máxima verosimilitud.

e) Discuta el resultado, prestando particular atención al rol del conocimiento que se posee sobre el modelo subyacente.

Sección 3.5

Ejercicio 4 El propósito de este problema es el derivar al clasificador Bayesiano desde el caso multivariado de Bernoulli. Como de costumbre, trabajamos con cada caso en forma separada, interpretando a $P(x|D)$ como la media $P(x|D_i, w_i)$. Sea la probabilidad condicional para una categoría dada definida por

$$P(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

y sea $D = \{x_1, \ldots, x_n\}$ un conjunto de n ejemplos seleccionados independientemente de acuerdo a esta densidad de probabilidad.

a) Si $s = (s_1, \ldots, s_d)^t$ es la suma de los n ejemplos, mostrar que

$$P(D|\theta) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i},$$

b) Asumiendo una distribución uniforme a priori para θ y usando la identidad

$$\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!},$$

muestre que

$$p(\theta|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i}(1-\theta_i)^{n-s_i},$$

c) Dibuje esta densidad para el caso $d = 1$, $n = 1$, y para los dos posibilidades resultados de s_1.

d) Integre el producto $P(x|\theta)p(\theta|D)$ sobre θ para obtener la probabilidad condicional deseada

$$P(x|D) = \prod_{i=1}^{d} \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}.$$

e) Si pensamos que obteniendo $P(x|D)$ sustituyendo un estimador $\hat{\theta}$ por θ en $P(x|\theta)$, cuál es el estimador Bayesiano efectivo para θ?

Sección 3.8

Ejercicio 5 Suponga que sa busca estimar θ describiendo una distribución multidimensional de datos D, algunos de cuyos puntos tienen características faltantes. Considere un algoritmo iterativo en el cual los valores de máxima verosimiltud en el cual son calculados los valores falantes, luego asuma que son correctos para la restitución de θ y la iteración.

a) Es esto equivalente a un algoritmo de "Expectation-Maximization", o sólo un algoritmo de "Expectation-Maximization" generalizado?

b) Si es un algoritmo de "Expectation-Maximization", qué es $Q(\theta, \theta^t)$, tal como se describe en la ecuación 78?

Sección 3.9

Ejercicio 6 Utilice las matrices de probabilidad condicional del ejemplo 3 para contestar los siguientes problemas sueltos:

a) Suponga que es 20 de Diciembre -el final del otoño y principio del invierno en el hemisferio norte- y por lo tanto sean $P(a_1) = P(a_2) = 0,5$. Además, se sabe que el pescado fue atrapado en el Atlántico Norte, es decir, $P(d_2) = 1$. Clasifique el pescado como Salmón o Mero. Cuál es el rango del error esperado?

b) Suponga que conocemos que el pescado es fino y con brillo medio. En qué estación estamos ahora, más probablemente? Cuál es la probabilidad de estar en lo cierto?

c) Suponga que conocemos que el pescado es fino y con brillo medio y que fue capturado en el Atlántico Norte. Cuál es la estación de año mas probable? Cuál es la probabilidad de estar en lo cierto?

Ejercicio 7 Una de las suposiciones más simples es la regla de Bayes ingenuo o la regla de Bayes tonta expresada en la ecuación 85. Dibuje la red de creencias para un problema de tres categorías con cinco características x_i, $i = 1, \ldots, 5$.

Ejercicio 8 Considere una red de creencias Bayesianas con varios nodos que tienen valores sin especificar. Suponga que uno de eos nodos es seleccionado al azar, las probabilidades de sus nodos son calculados por las fórmulas descriptas en el texto. Luego otro nodo es seleccionado al azar (posiblemente, incluso un nodo ya visitado), y las probabilidades actualizadas en forma similar. Pruebe que este procedimiento converge a las probabilidades deseadas a lo largo de toda la red.

Sección 3.10

Ejercicio 9 Elmétodo estándar para el cálculo de probabilidad de una secuencia en una Hidden Markov Model dada, es a través del uso de probabilidades a posteriori $\alpha_i(t)$.

a) Muestre por sustitución simple que un método simétrico puede ser derivado usando las probabilidades hacia atrás $\beta_i(t)$.

b) Pruebe que se puede obtener la probabilidad por combinación de probabilidades hacia adelante y hacia atrás en cualquier lugar del medio de la secuencia. Esto es, muestre que

$$P(w^{T'}) = \sum_{i=1}^{T'} \alpha_i(t)\beta_i(t),$$

donde $w^{T'}$ es una secuencia particular de longitud $T' < T$.

c) Muestre que su fórmula se reduce al conocimiento de los valores al principio y al final de la secuencia.

Ejercicio 10 Explore la relación cercana entre redes de creencia Bayesianas y el Hidden Markov Model de la siguiente forma:

a) Pruebe que las ecuaciones hacia adelante y hacia atrás para Hidden Markov Model son casos especiales de Eq. 84.

b) Use su respuesta para explicar la relación entre estos dos tipos generales de modelos.

Ejercicios Computacionales

Varios ejercicios harán uso de los siguientes datos tridimensionales de la muestra a partir de tres categorías, denotado w_i.

Table 1: Tabla de Datos

	w_1			w_2			w_3		
Punto	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	0.42	-0.087	0.58	-0.4	0.58	0.089	0.83	1.6	-0.014
2	-0.2	-3.3	-3.4	-0.31	0.27	-0.04	1.1	1.6	0.48
3	1.3	-0.32	1.7	0.38	0.055	-0.035	-0.44	-0.41	0.32
4	0.39	0.71	0.23	-0.15	0.53	0.011	0.047	-0.45	1.4
5	-1.6	-5.3	-0.15	-0.35	0.47	0.034	0.28	0.35	3.1
6	-0.029	0.89	-4.7	0.17	0.69	0.1	-0.39	-0.48	0.11
7	-0.23	1.9	2.2	-0.011	0.55	-0.18	0.34	-0.079	0.14
8	0.27	-0.3	-0.87	-0.27	0.61	0.12	-0.3	-0.22	2.2
9	-1.9	0.76	-2.1	-0.065	0.49	0.0012	1.1	1.2	-0.46
10	0.87	-1.0	-2.6	-0.12	0.054	-0.063	0.18	-0.11	-0.49

Sección 3.2

Ejercicio 1 Considere los modelos de densidad Gaussiana en diferentes dimensiones:

a) Escriba un programa para encontrar los valores de máxima verosimilitud $\hat{\mu}$ y $\hat{\sigma}^2$. Aplique su programa individualmente para cada una de las tres característica x_i de la categoría w_1 de la *Table 1*.

b) Modifique su programa para aplicarlo a datos Gaussianos bidimensionales $p(x) \sim N(\mu, \Sigma)$. Aplique su programa para cadas una de las tres formas de apareamiento de a dos características para w_1.

c) Modifique su programa para aplicarlo a datos Gaussianos de tres dimensiones. Aplique su programa para a las tres dimensiones de la categoría w_1.

d) Asuma que su modelo de tres dimensiones es separable, por lo tanto $\Sigma = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Escriba un programa para estimar la media y los componentes diagonales de Σ. Aplique el programa a los datos en w_2.

e) Compare los resultados para la media de cada característica μ_i calculada en las formas previas. Explique por que son iguales o diferentes.

f) Compare sus resultados para la varianza de cada característica σ_i^2 calculada de las formas previas. Explique por que los resultados son iguales o diferentes.

Sección 3.3

Ejercicio 2 Considere un modelo unidimensional de una densidad triangular gobernada por dos parámetros escalares:

$$p(x|\theta) \equiv T(\mu,\delta) = \begin{cases} \frac{(\delta-|x-\mu|)}{\delta^2} & \text{para } |x-\mu| < \delta \\ 0 & \text{en caso contrario} \end{cases}$$

donde $\theta = \binom{\mu}{\delta}$. Escriba un programa para calcular la densidad $p(x|D)$ a través de métodos Bayesianos (Eq. 26) y aplicarlos a la característica x_2 de la catagoría w_2. Grafique la densidad a posteriori resultante $p(x|D)$.

Sección 3.7

Ejercicio 3 Considerar las tasas de error en diferentes dimensiones

a) Use máxima verosimilitud para entrenar un dicotomizador usando los datos tridimensionales de las categorías w_1 y w_2 en la *Table 1*. Integre numéricamente para estimar la proporción del error.

b) Ahora considere los datos proyectados en un subespacio bidimensional. Para cada uno de los tres subespacios - definidos por $x_1 = 0$, $x_2 = 0$ o $x_3 = 0$- entrene un dicotomizador Gaussiano. Integre numéricamente para estimar la proporción del error.

c) Ahora considere los datos proyectados en subespacios unidimensionales, definidos por cada uno de los tres ejes. Entrene un clasificador Gaussiano e integre numéricamente para estimar la proporción del error.

d) Discuta el orden del rango de las tasas de error encontrado.

e) Asumiendo que reestima la distribución en diferentes dimensiones, lógicamente el error de Bayes debe ser mayor en los espacios proyectados.

Sección 3.9

Ejercicio 4 Escriba un programa para evaluar una red de creencias Bayesianas para el pez en el ejemplo 3, incluyendo la información en $P(x_i|a_j)$, $P(x_i|b_j)$, $P(c_i|x_j)$ y $P(d_i|x_j)$. Verifique su programa sobre los cálculos dados en el ejemplo. Aplique su programa en los siguientes casos y consigne toda suposición que necesite realizar.

a) Un pez oscuro y fino es capturado en el Atlántico Norte en verano. Cuál es la probabilidad de que sea un salmón?

b) Un pez de brillo mediano y fino es capturado en el Atlántico Norte. Cuál es la probabilidad de que sea invierno, primavera, verano u otoño?

c) Un pez claro y ancho es capturado en otoño. Cuál es la probabilidad de que venga del Atlántico Norte?

Sección 3.10

Ejercicio 5 Considere el uso de Hidden Markov Model para clasificar secuencias de cuatro estados visibles, A-D. Entrene dos Hidden Markov Model, cada uno con tres estados ocultos (además de un estado nulo inicial y otro estado nulo final), totalmente conectado, con los datos de la *Table 2*. Asuma que cada secuencia inicia con un símbolo nulo y termina con símbolo nulo final (no figuran).

Table 2: Tabla de Datos

Ejemplo	w_1	w_2
1	AABBCCDD	DDCCBBAA
2	ABBCBBDD	DDABCBA
3	ACBCBCD	CDCDCBABA
4	AD	DDBBA
5	ACBCBABCDD	DADACBBAA
6	BABAADDD	CDDCCBA
7	BABCDCC	BDDBCAAAA
8	ABDBBCCDD	BBABBDDDCD
9	ABAAACDCCD	DDADDBCAA
10	ABD	DDCAAA

a) Imprima las matrices de transición completas para cada uno de los modelos.

b) Asuma probabilidades a priori iguales para los dos modelos y clasifique cada una de las siguientes secuencias: ABBBCDDD, DADBCBAA, CDCBABA y ADBBBCD.

c) Como antes, clasifique el patrón de prueba BADBDCBA. Encuentre las prioridades a priori de sus dos modelos entrenados que dejarían iguales probabilidades a posteriori para sus dos categorías cuando se aplican a este patrón.