Notación

- P_{FA}: probabilidad de falsa alarma
- ML: máxima verosimilitud
- $P(i|x)$: probabilidad, *a posteriori*, para la hipótesis $H = i$ dado un valor de x

1. Sean X_1 y X_2 dos variables aleatorias independientes. X_1 sigue una distribución exponencial unilateral:

$$p_{X_1}(x_1) = \frac{1}{\theta} \exp\left(-\frac{x_1}{\theta}\right) \qquad 0 \leq x_1 < \infty$$

donde $\theta > 0$ es un parámetro determinista de valor desconocido, mientras que X_2 sigue una distribución uniforme en el intervalo $[0, \theta]$, representando θ el mismo parámetro en ambas distribuciones.

Para estimar el valor de θ, se propone el siguiente estimador:

$$\widehat{\theta} = aX_1 + bX_2$$

donde a y b son dos constantes.

(a) Determine la esperanza matemática (valor medio) y varianza de la variable aleatoria X_1.

(b) Determine la esperanza matemática (valor medio) y varianza de la variable aleatoria X_2.

(c) ¿Es el estimador $\widehat{\theta}$ insesgado? Justifique su respuesta.

(d) Calcule la varianza del estimador $\widehat{\theta}$.

(e) Obtenga el valor de la constante b, en función de a, que hace que el estimador $\widehat{\theta}$ sea insesgado. Reescriba la expresión del estimador $\widehat{\theta}$ para que dependa sólo de la constante a.

(f) Determine el valor de a, tal que el estimador obtenido en el apartado anterior tenga la menor varianza posible. ¿Cuál sería, por tanto, la expresión del estimador $\widehat{\theta}$ insesgado de mínima varianza?

Solution:

(a) $\mathbb{E}\{X_1\} = \theta$; $\mathrm{Var}\{X_1\} = \theta^2$.

(b) $\mathbb{E}\{X_2\} = \dfrac{\theta}{2}$; $\mathrm{Var}\{X_2\} = \dfrac{\theta^2}{12}$.

(c) $\mathrm{Sesgo}(\widehat{\theta}) = \left(a + \dfrac{b}{2} - 1\right)\theta$. El estimador $\widehat{\theta}$ es sesgado.

(d) $\mathrm{Var}(\widehat{\theta}) = \left(a^2 + \dfrac{b^2}{12}\right)\theta^2$

(e) $b = 2(1-a)$ $\widehat{\theta} = aX_1 + 2(1-a)X_2$

(f) $a^* = \dfrac{1}{4}$

$\widehat{\theta}^* = \dfrac{X_1}{4} + \dfrac{3X_2}{2}$.

(Puntuación: 1,5/4 puntos)

2. Un sistema de vigilancia se encarga de detectar intrusos en un edificio. Se considera el siguiente par de hipótesis:

   - $H = 0$: no hay intruso
   - $H = 1$: hay un intruso presente

   El sistema envía un mensaje de alarma si se acepta la hipótesis $H = 1$. Suponga que después de procesar los datos, se obtiene $P(1|x) = 0.05$. Asumiendo que el coste de no detectar un intruso cuando sí lo hay es 10 veces el coste de decidir que hay intruso cuando no es cierto, ¿debería el sistema enviar un mensaje de alarma (es decir, aceptar la hipótesis $H = 1$)? Justifique su respuesta.

   Nota: para la resolución del problema, considere que el coste de acierto asociado a cada hipótesis es nulo.

Solution:

Dado que $P(0|x)\, c_{10} > P(1|x)\, c_{01}$, se acepta la hipótesis $H = 0$, y por tanto, **no es necesario** enviar un **mensaje de alarma**.

(Puntuación: 0,75/4 puntos)

3. Sabiendo que N es una variable aleatoria cuya función de densidad de probabilidad es una gaussiana de media 0 y varianza 1:

$$p_N(n) = \mathrm{G}\big(n|0, 1\big)$$

Considere el par de hipótesis:

$$\begin{aligned} H = 0: &\quad X = \sqrt{2}N \\ H = 1: &\quad X = 1 + N \end{aligned}$$

(a) Determine $p_{X|H}(x|0)$ y $p_{X|H}(x|1)$. Represente, en la misma gráfica, ambas verosimilitudes.

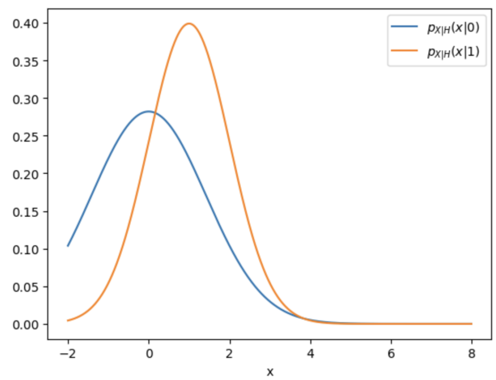
(b) Obtenga el clasificador ML basado en X.

(c) Calcule la probabilidad de falsa alarma del clasificador ML. Exprese esta probabilidad utilizando la función:

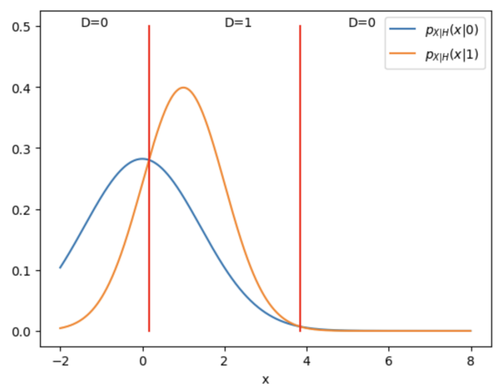
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

Solution:

(a) $p_{X|H}(x|0) = \mathrm{G}(0, 2)$
$p_{X|H}(x|1) = \mathrm{G}(1, 1)$



(b) $\phi_{\mathrm{ML}} = \begin{cases} D=0 & x < 0.16 \ \cup \ x > 3.84 \\ D=1 & 0.16 < x < 3.84 \end{cases}$



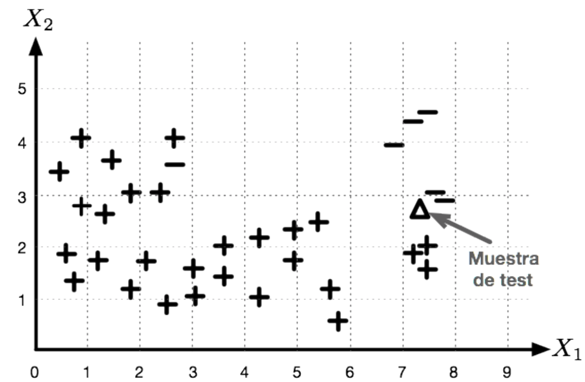
(c) $P_{\mathrm{FA}} = F\left(\frac{3.84}{\sqrt{2}}\right) - F\left(\frac{0.16}{\sqrt{2}}\right)$

(Puntuación: 0,75/4 puntos)

Notación

- k-NN: del inglés, k-*nearest neighbors*. Algoritmo de los k vecinos más cercanos.

4. Suponga que se dispone del siguiente conjunto de datos de entramiento en el que se incluyen muestras de la clase positiva (+) y muestras de la clase negativa (−), así como una muestra de test (△).



Todas las muestras se proyectan en un espacio vectorial de 2 características (X_1 y X_2). Suponiendo que se utiliza el algoritmo de clasificación k-NN ponderado, responda a las siguientes preguntas:

(a) ¿Cuál sería la clase asignada para la muestra de test si $k = 1$? Razone su respuesta.

(b) Para este conjunto de datos de entrenamiento, ¿recomendaría utilizar $k = 11$? Razone su respuesta.

Solution:

(a) Para $k = 1$, la muestra de test sería clasificada como clase negativa (−) dado que su vecino más próximo del conjunto de entrenamiento está etiquetado con la clase negativa (cuando $k = 1$, el algoritmo k-NN ponderado se comporta como el algoritmo k-NN uniforme).

(b) Dado que sólo hay 5 muestras etiquetadas con la clase negativa en el conjunto de datos de entrenamiento, para cualquier valor de $k > 10$, en el algoritmo k-NN uniforme, la muestra de test sería etiquetada con la clase positiva. Sin embargo, si se usa el algoritmo k-NN ponderado, dado que se da más peso a los vecinos más próximos, para este conjunto de datos, a la muestra de test mostrada en la gráfica, se le asignaría la clase negativa (−), así que sí recomendaría utilizar $k = 11$.

(Puntuación: 0,3/4 puntos)

5. Razone si la siguiente afirmación es cierta o falsa:

"Para el regresor Lasso, si el parámetro de regularización es cero, la función de pérdida que se minimiza para obtener los coeficientes es la suma de los cuadrados de los errores entre los valores reales de las etiquetas y los valores predichos".

En caso de ser falsa, haga la(s) correccione(s) necesaria(s) para que la afirmación sea cierta.

Solution: La afirmación es cierta.
En regresión Lasso, se introduce un término de regulizarización en la función de pérdida (o de coste), de manera que se minimice, en este caso, también la norma L1 del vector de coeficientes:

$$\min_{w_0,\mathbf{w}} \frac{1}{N}\sum_{i=1}^{N}(y_i - (w_0 + \mathbf{w}^\top \mathbf{x}))^2 + \alpha \sum_{j=1}^{d} |w_j|$$

donde α es el parámetro (o hiperparámetro) de regularización. Si $\alpha = 0$, la función de pérdida es la suma de los cuadrados de los errores entre los valores reales de las etiquetas y los valores predichos (o viceversa).

(Puntuación: 0,3/4 puntos)

6. Responda brevemente a las siguientes preguntas:
   1. ¿Es el filtro de Wiener un filtro lineal? Justifique su respuesta.
   2. ¿Qué es un filtro adaptativo? Indique un ejemplo de tipo de algoritmo que se usa en filtros adaptativos para encontrar los coeficientes del mismo.

Solution:

1. Sí, lo es. La salida de un filtro de Wiener de orden P se calcula como una combinación lineal del vector de entrada y los coeficientes del filtro.

$$y[n] = \sum_{k=0}^{P-1} w[k]x[n-k]$$

   donde:

   - $\mathbf{w}$ es el vector con los P coeficientes del filtro.

$$\mathbf{w}^\top = \begin{bmatrix} w[0] & w[1] & \cdots & w[P-2] & w[P-1] \end{bmatrix}$$

   - $\mathbf{x}[n]$ es un vector con P elementos resultantes de la concatenación de la muestre n-ésima y las $P-1$ muestras correspondientes a los instantes de tiempo $P-1$ anteriores a n:

$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-P+2] \\ x[n-P+1] \end{bmatrix}.$$

2. En el diseño de un filtro adaptativo (variante en el tiempo), para cada instante de tiempo n, se ha de encontrar un conjunto de coeficientes óptimos, $\mathbf{w}_k(n)$ para $k = 0, 1, \ldots, P-1$. El problema se simplifica considerando una actualización de los pesos del filtro de la forma:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \nabla \mathbf{w}_n$$

   donde $\nabla \mathbf{w}_n$ es un factor de corrección que se aplica a los coeficientes $\mathbf{w}(n)$ en el instante n para obtener el nuevo conjunto de coeficientes $\mathbf{w}(n+1)$ en el instante

$n+1$. Esta ecuación de actualización es la base de los algoritmos adaptativos, y el diseño de cada filtro adaptativo requiere definir este factor de correción $\nabla \mathbf{w}_n$.

Uno de los algoritmos más conocidos es el algoritmo LMS (del inglés, *Least-Mean-Square*) debido a que su bajo coste computacional permite su implementación en una gran variedad de aplicaciones.

(Puntuación: 0,4/4 puntos)