1 Teorico

(b) El metodo de validacion cruzada tiene como funcion objetivo a:

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{m_h}^{-i}(X_i))^2$$

para la minimizacion en h, donde el estimador $\widehat{m_h}^{-i}(.)$ se computa de la siguiente manera:

$$\widehat{m_h}^{-i}(x) = \sum_{j=1, j\neq i}^{n} Y_j \frac{g_j(x)}{\sum\limits_{t=1,t\neq i}^{n} g_t(x)}$$

Donde, para mayor legibilidad en las cuentas que vendran, escribimos

$$g_j(x) = K\left(\frac{X_j - x}{h}\right)$$

Notar que el estimador $\widehat{m_h}^{-i}(.)$ se computa como en la igualdad (1) del enunciado, pero sin la i-esima observacion, (X_i, Y_i). En todo lo que sigue, tomaremos un $i \leq n$ fijo. Dado que este i es arbitrario, lo siguiente que probaremos sera valido para la validacion cruzada sacando cualquiera de las observaciones.

Queremos probar la igualdad (2):

$$\widehat{m_h}^{-i}(X_i) = \frac{\widehat{m_h}(X_i) - Y_i w_{i,h}(X_i)}{1 - w_{i,h}(X_i)}$$

Donde, en concordancia con la notacion del enunciado, llamamos:

$$w_{i,h}(x) = \frac{g_i(x)}{\sum\limits_{t=1}^{n} g_t(x)}$$

Para probar (2), notemos que tenemos a $\widehat{m_h}^{-i}(x)$ escrito como una sumatoria sobre $j = 1, 2, .., n$ con $j \neq i$. Estamos sumando Y_j mutiplicado por una fraccion, analizamos esa fraccion:

$$\frac{g_j(x)}{\sum\limits_{t=1,t\neq i}^{n} g_t(x)} = \frac{g_j(x)}{-g_i(x) + \sum\limits_{t=1}^{n} g_t(x)}$$

Multiplicamos por un 1 convenientemente escrito, osea, $\sum\limits_{t=1}^{n} g_t(x)$ sobre si mismo:

$$\frac{g_j(x)}{-g_i(x)+\sum\limits_{t=1}^{n} g_t(x)} = \frac{g_j(x)}{\sum\limits_{t=1}^{n} g_t(x)} * \frac{\sum\limits_{t=1}^{n} g_t(x)}{-g_i(x)+\sum\limits_{t=1}^{n} g_t(x)} = w_{j,h}(x) * \frac{\sum\limits_{t=1}^{n} g_t(x)}{-g_i(x)+\sum\limits_{t=1}^{n} g_t(x)}$$

Notar que el segundo multiplicando no depende de j. Reescribamos a este multiplicando con una expresion mas amigable:

$$\frac{\sum\limits_{t=1}^{n} g_t(x)}{-g_i(x)+\sum\limits_{t=1}^{n} g_t(x)} = \left(\frac{-g_i(x)+\sum\limits_{t=1}^{n} g_t(x)}{\sum\limits_{t=1}^{n} g_t(x)}\right)^{-1} = \left(-\frac{g_i(x)}{\sum\limits_{t=1}^{n} g_t(x)} + \frac{\sum\limits_{t=1}^{n} g_t(x)}{\sum\limits_{t=1}^{n} g_t(x)}\right)^{-1} = (-w_{i,h}(x)+1)^{-1}$$

Ahora si, con esta informacion, podemos reescribir a $\widehat{m_h}^{-i}(.)$ como:

$$\widehat{m_h}^{-i}(x) = \sum_{j=1, j\neq i}^{n} Y_j \frac{g_j(x)}{\sum\limits_{t=1, t\neq i}^{n} g_t(x)} = \sum_{j=1, j\neq i}^{n} Y_j \frac{w_{j,h}(x)}{1-w_{i,h}(x)}$$

Sacamos convenientemente el divisor de la fraccion, que no depende de j, de la sumatoria:

$$\widehat{m_h}^{-i}(x) = \frac{1}{1-w_{i,h}(x)} \sum_{j=1, j\neq i}^{n} Y_j * w_{j,h}(x)$$

Sabiendo que $\widehat{m_h}(x) = \sum_{j=1}^{n} Y_j * w_{j,h}(x)$, podemos reescribir la sumatoria de arriba como

$$\widehat{m_h}(x) - Y_i * w_{i,h}(x) = \sum_{j=1, j\neq i}^{n} Y_j * w_{j,h}(x)$$

Finalmente, juntando todo lo recien visto y evaluando $\widehat{m_h}^{-i}(x)$ en X_i, probamos la igualdad (2)

$$\widehat{m_h}^{-i}(X_i) = \frac{\widehat{m_h}(X_i) - Y_i * w_{i,h}(X_i)}{1-w_{i,h}(X_i)}$$

Para probar la igualdad (3), que establece que

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n} \frac{(Y_i - \widehat{m_h}(X_i))^2}{(1 - w_{i,h}(X_i))^2}$$

reemplazaremos lo obtenido en (2) en la definicion de $CV(h)$ que enunciamos al principio.

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{m_h}^{-i}(X_i))^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{\widehat{m_h}(X_i) - Y_i * w_{i,h}(X_i)}{1 - w_{i,h}(X_i)}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i * (1 - w_{i,h}(X_i))}{1 - w_{i,h}(X_i)} - \frac{\widehat{m_h}(X_i) - Y_i * w_{i,h}(X_i)}{1 - w_{i,h}(X_i)}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - Y_i * w_{i,h}(X_i))}{1 - w_{i,h}(X_i)} - \frac{\widehat{m_h}(X_i) - Y_i * w_{i,h}(X_i)}{1 - w_{i,h}(X_i)}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - Y_i * w_{i,h}(X_i)) - \widehat{m_h}(X_i) + Y_i * w_{i,h}(X_i)}{1 - w_{i,h}(X_i)}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{m_h}(X_i)}{1 - w_{i,h}(X_i)}\right)^2$$

Que es lo que queriamos probar.