Note 1:

$y = f(x) + \xi$, $E(\xi) = 0$ & $Var(\xi) = \sigma_\xi^2$.

$f(x)$: is not random but unknown.

$$Err(x_0) = E\{(y - \hat{f}(x_0))^2\}$$

$$= E\{(f(x_0) + \xi - \hat{f}(x_0))^2\}$$

$$= E\{(f(x_0) - \hat{f}(x_0))^2 + 2\xi(f(x_0) - \hat{f}(x_0)) + \xi^2\}$$

$$= E\{(f(x_0) - \hat{f}(x_0))^2\} + 2E\{\xi(f(x_0) - \hat{f}(x_0))\} + E\{\xi^2\}$$

By independence of the error (and helped if $\hat{f}$ is unbiased)

$$E\{\xi(f(x_0) - \hat{f}(x_0))\} = E(\xi)\,E\{f(x_0) - \hat{f}(x_0)\} = 0$$

$\hat{f}$ is random as it is estimated from y.

$$E\{[f(x_0) - E\hat{f}(x_0) + E\hat{f}(x_0) - \hat{f}(x_0)]^2\} = E\{[f(x_0) - \hat{f}(x_0)]^2\}$$

$$= E\{[f(x_0) - E\hat{f}(x_0)]^2 + [E\hat{f}(x_0) - \hat{f}(x_0)]^2 +$$
$$2E(f(x_0) - E\hat{f}(x_0))(E\hat{f}(x_0) - \hat{f}(x_0))\}$$

$f(x_0)$ & $E\hat{f}(x_0)$ are not random

$$E[E\hat{f}(x_0) - \hat{f}(x_0)] = E\hat{f}(x_0) - E\hat{f}(x_0) = 0$$

$$\Rightarrow 2E\{(f(x_0) - E\hat{f}(x_0))(E\hat{f}(x_0) - \hat{f}(x_0))\} = 0$$

$$Err(x_0) = [f(x_0) - E\hat{f}(x_0)]^2 + E[(E\hat{f}(x_0) - \hat{f}(x_0))^2] + \sigma_\xi^2.$$

Note 2:

$$\hat{f}(x_i) = x_i^T \hat{\beta} = x_i^T (X^T X) X^T y = x_i^T (X^T X)^{-1} X^T (X\beta + \xi)$$

$$= x_i^T (X^T X)^{-1} (X^T X)\beta + x_i^T (X^T X)^{-1} X^T \xi$$

$$= x_i^T \beta + x_i^T (X^T X)^{-1} X^T \xi$$

$$Var\, \hat{f}(x_i) = x_i^T (X^T X)^{-1} X^T Var(\xi) X (X^T X)^{-1} x_i$$

$$= \sigma_\xi^2 x_i^T (X^T X)^{-1} x_i = \sigma_\xi^2 \|h(x_i)\|^2$$

$$\sum_{i=1}^{N} x_i^T (X^T X)^{-1} x_i = \sum_{i=1}^{N} trace[x_i^T (X^T X)^{-1} x_i]$$

$$= \sum_{i=1}^{N} trace[x_i x_i^T (X^T X)^{-1}]$$

$$= trace \sum_{i=1}^{N} x_i x_i^T (X^T X)^{-1}$$

$$= trace (X^T X)(X^T X)^{-1} = trace\, I_p = p.$$

Note 3:

$$Err = \frac{1}{N} \sum_{i=1}^{N} E_y[(y_i - \hat{f}(x_i))^2]$$

$$\overline{err} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{f}(x_i))^2$$

$$y_i - \hat{f}(x_i) = y_i - f(x_i) + f(x_i) - E\hat{f}(x_i) + E\hat{f}(x_i) - \hat{f}(x_i)$$

$$(y_i - \hat{f}(x_i))^2 = A_1 + B + C + D_1 + E + F_1$$

$$\sum_i A_1 = \sum_i (y_i - f(x_i))^2$$

$$\sum_i B = \sum_i (f(x_i) - E\hat{f}(x_i))^2$$

$$\sum_i C = \sum_i (E\hat{f}(x_i) - \hat{f}(x_i))^2$$

$$\sum_i D_1 = 2\sum_i (y_i - f(x_i))(f(x_i) - E\hat{f}(x_i))$$

$$\sum_i E = 2\sum_i (f(x_i) - E\hat{f}(x_i))(E\hat{f}(x_i) - \hat{f}(x_i))$$

$$\sum_i F_1 = 2\sum_i (y_i - f(x_i))(E\hat{f}(x_i) - \hat{f}(x_i))$$

$$E_y(y_i - \hat{f}(x_i))^2 = A_2 + B + C + D_2 + E + F_2$$

$$\sum_i A_2 = \sum_i E_y(y_i - f(x_i))^2$$

$$\sum_i D_2 = 2\sum_i E_y(y_i - f(x_i))(f(x_i) - E\hat{f}(x_i))$$

$$\sum_i F_2 = 2\sum_i E_y(y_i - f(x_i))(E\hat{f}(x_i) - \hat{f}(x_i))$$

$$N[Err - \overline{err}] = \sum_i (A_2 + B + C + D_2 + E + F_2) - (A_1 + B + C + D_1 + E + F_1)$$

$$= \sum_i (A_2 - A_1) + (D_2 - D_1) + (F_2 - F_1)$$

$$E(A_1) = E(A_2) = N\sigma_\varepsilon^2 \rightarrow \text{unpredictable error variance}$$

$$\Rightarrow E(A_2 - A_1) = 0$$

$$\sum_i E(D_1) = 2\sum_i (E(y_i) - f(x_i))(f(x_i) - E\hat{f}(x_i)) = 0 \text{ since } E(y_i) = f(x_i)$$

$E(D_2) = 0$ for similar reasons

$$\sum_i F_2 = 2\sum_i E_y (y_i - f(x_i))(\hat{f}(x_i) - \hat{f}(x_i)) = 0$$

since y_i is independent of $\hat{f}(x_i)$ and $E_y(y_i) = f(x_i)$

$$\sum_i E(F_1) = 2\sum_i E\left[(y_i - f(x_i))(E\hat{f}(x_i) - \hat{f}(x_i))\right]$$

Note that: $E(y_i) = f(x_i)$ & $E(\hat{f}(x_i)) = E\hat{f}(x_i)$

$$\begin{cases} y_i - f(x_i) \\ \hat{f}(x_i) - E\hat{f}(x_i) \end{cases}$$ Terms that are the difference between the sample and the expectation

$$\sum_i E(F_1) = -2\sum_{i=1}^{N} \text{cov}(y_i, \hat{y}_i)$$

$$E\left[N(Err - \overline{err})\right] = \sum_{i} E(F_i) = 2\sum_{i=1}^{N} \operatorname{cov}(y_i, \hat{y}_i)$$

Note 4:

$$\therefore \hat{y}_i = e_i^T X (X^T X)^{-1} X^T y$$

position "i"

$$y_i = e_i^T y, \quad e_i^T = (0, 0 \ldots, 1, 0, \ldots 0)$$

$$\operatorname{cov}(\hat{y}_i, y_i) = e_i^T \operatorname{cov}(\hat{y}, y)\, e_i$$

$$\operatorname{cov}(\hat{y}, y) = X (X^T X)^{-1} X^T \operatorname{cov}(y, y) = \sigma^2 X (X^T X)^{-1} X^T$$

$$\operatorname{cov}(\hat{y}_i, y_i) = \sigma^2 (e_i^T X)(X^T X)^{-1} (X^T e_i)$$

$$= \sigma^2 x_i^T (X^T X)^{-1} x_i$$

$$\sum_{i=1}^{N} x_i^T (X^T X)^{-1} x_i = d$$

$$\sum_{i=1}^{N} \operatorname{cov}(\hat{y}_i, y_i) = \sigma^2 d$$

Note 5:

$X = (x_1, \dots, x_n)'$ and $y = (y_1, \dots, y_n)'$

$$\hat{\beta} = (X'X)^{-1} X'y$$

if we remove (x_i', y_i) then

$$\hat{\beta}_{(i)} = (X'X - x_i x_i')^{-1} X'y - x_i y_i$$

using the identity: $(A - bb')^{-1} = A^{-1} + A^{-1} bb' A^{-1}/d$ where $d = 1 - b'A^{-1}b$ is a scalar

$$\hat{\beta}_i = \left[(X'X)^{-1} + (X'X)^{-1} x_i x_i' (X'X)^{-1}/(1 - s_{ii})\right](X'y - x_i y_i)$$

where: $s_{ii} = x_i'(X'X)^{-1} x_i$

$$y_i - \hat{f}^{-i}(x_i) = y_i - x_i'\left[(X'X)^{-1} + (X'X)^{-1} x_i x_i' (X'X)^{-1}/(1 - s_{ii})\right](X'y - x_i y_i)$$

$$= y_i - \left[1 + x_i'(X'X)^{-1} x_i/(1 - s_{ii})\right] x_i'(X'X)^{-1}(X'y - x_i y_i)$$

$$= y_i - \left[1 + \frac{s_{ii}}{1 - s_{ii}}\right] x_i'(X'X)^{-1}(X'y - x_i y_i)$$

$$= y_i - \frac{1}{1 - s_{ii}} x_i'(X'X)^{-1}(X'y - x_i y_i)$$

$$= y_i\left(1 + \frac{x_i'(X'X)^{-1}x_i}{1 - s_{ii}}\right) - \frac{1}{1 - s_{ii}} x_i'\hat{\beta} = y_i\left[1 + \frac{s_{ii}}{1 - s_{ii}}\right] - \frac{1}{1 - s_{ii}} x_i'\hat{\beta}$$

$$= (y_i - x_i'\hat{\beta})/(1 - s_{ii}) = \frac{y_i - \hat{f}(x_i)}{1 - s_{ii}}$$

Note 6:

For $\hat{y} = Sy$

$$GCV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \hat{f}(x_i)}{1 - trace(S)/N}\right]^2$$

Using the approximation $\frac{1}{(1-x)^2} \approx 1 + 2x$ we have

$$GCV(\hat{f}) \approx \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2\left(1 + \frac{2\,trace(S)}{N}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2 + \frac{2}{N^2}\,trace(S)\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2$$

Taking

$$\hat{\sigma}_{\varepsilon}^2 \approx \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2$$

$$GCV(\hat{f}) = \overline{err} + \frac{2\,trace(S)}{N}\hat{\sigma}_{\varepsilon}^2$$

In the case of linear regression: $trace(S) = p$

$$GCV(\hat{f}) = \overline{err} + \frac{2p}{N}\hat{\sigma}_{\varepsilon}^2$$