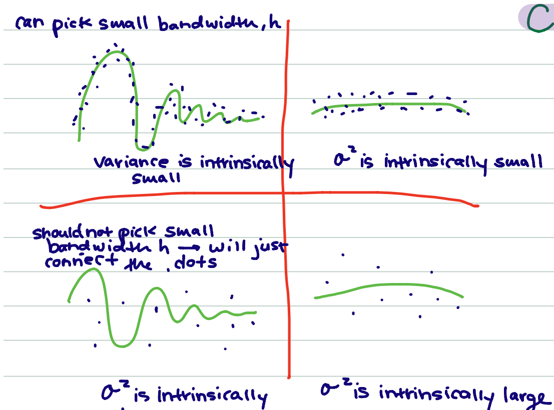
Cross Validation



- need to be careful in choosing bandwidth when picking h

50% "train"
50% "test"

200 data points ⟹ 200 d.p.

fit → 100, h = bandwidth that's optimal for fitting estimator (but what abt 200 data pts? → high bias)

predict → 100, h

fit → 199, h: trying to estimate generalization error; make bias negligible but blow up σ^2

predict → 1, h

∴ cross validation estimate is subject to error regardless of strategy
"no free lunch" → no panacea for choosing hyperparameters

Exercise 2

c) LOOCV = "Leave One Out Cross Validation"

$$H = \left(\frac{W(x_i,x_j)}{\sum_k (x_i,x_k)}\right) \qquad \hat{y} = Hy \qquad \hat{y}_i^{(-i)} = \frac{\left(\sum w(x_i,x_j)\, y_j - w(x_i,x_i)\, y_i\right.}{\sum (w(x_i,x_j)) - w(x_i,x_i)} \qquad w_{ij} = w(x_i,x_j)$$

$$\hat{y}_i^{(-i)} = \frac{\overbrace{H_i y}^{\hat{y}_i} - H_{ii} y}{1-H_{ii}} = (\hat{y} - H_{ii} y_i)\left(\frac{1}{1-H_{ii}}\right) \Rightarrow \hat{y}_i^{(-i)} = \hat{y}_i - H_{ii} y_i + H_{ii} \hat{y}_i^{(-i)}$$

$$\text{LOOCV} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i^{(-i)}\right)^2$$

Remember: $\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty = Hy$

$$= \left(y_i - (\hat{y}_i - H_{ii} y_i)\frac{1}{1-H_{ii}}\right)^2 = y_i - \frac{H_{ii}}{1-H_{ii}} y_i - \frac{\hat{y}_i}{1-H_i} - \left(\left(1 - \frac{H_{ii}}{1-H_{ii}}\right) y_i - \frac{\hat{y}}{1-H_{ii}}\right)^2$$

$$= \sum \left(\frac{y_i - \hat{y}_i}{1-H_{ii}}\right)^2$$

* this does not work for Least Squares

Another method:

$\hat{y} = Hy \qquad \hat{y}_i = \Sigma_j H_{ij} y_j$

Let $Z_j = \begin{cases} y_j & j \neq i \\ y_j^{(-i)} & j = i \end{cases}$

$y_j^{(-i)} = \Sigma_j H_{ij} Z_j$

$\hat{y}_i - \hat{y}_i^{(-i)} = \Sigma H_{ij} (y_j - Z_j) = H_{ii} (y_i - \hat{y}_i^{(-i)})$

Local Polynomial Regression

A) $g_x(x_i; a) = a_0 + R_{x_i} a \quad [(x-x_i)(x-x_i)^2 \ldots (x-x_i)^n]$ $\qquad R_x \begin{bmatrix} R_{x_1} \\ \vdots \\ R_{x_n} \end{bmatrix}$

$\hat{a} = [a_0 \; a] = \underset{a}{\operatorname{argmin}} \sum_i w_i (y_i - a_0 - R_{x_i} a)^2$

$a_0 = \dfrac{w^T (y - R_x a)}{w^T 1}$ $\qquad \tilde{y} = y - a_0 1$

$\dfrac{\partial}{\partial a} \left[(\tilde{y} - R_x a \ldots \right.$

$R(f, \hat{f}_n) = \int_0^1 \overset{\text{risk}}{R}[f(x), \hat{f}_n(x)]\, dx$

$R(f, \hat{f}_n^{(N-W)}) = \dfrac{h_n}{4} \left(\int \overset{\text{kernel}}{x^2 k(x) dx} \right) \int \left(\overset{\text{true function}}{f''(x) + 2f'(x) \dfrac{\pi'(x)}{\pi(x)}} \right)^2 dx + \overset{\text{variance}}{\sigma^2 \dfrac{\int k^2(x)}{n h_n}} + \ldots o(n h_n^{-1})$

$\pi(x)$ = density of x's