Lecture November 8

In ML we deal with samples of data

$$E[y] = \frac{1}{n} \sum_{i=0}^{n-1} y_i = \bar{\mu}_y$$

(sample mean)

$$\mathrm{var}[y] = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \bar{\mu}_y)^2$$

True mean $\mu_y \neq \bar{\mu}_y$

In order to estimate $\bar{\mu}_y$ and $\mathrm{var}[y]$, we use resampling methods

- Bootstrap
- Cross-validation

Bootstrap

$$D = \{x_0, x_1 \dots x_{n-1}\}$$

(i) select new $D^{(1)}$ by selecting randomly (with replacement) n-point in D

$$D^{(1)} = \{x_0^*, x_1^*, \ldots x_{n-1}^*\}$$

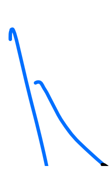
(ii) compute $\mu_x^{(1)}$, $MSE^{(1)}$

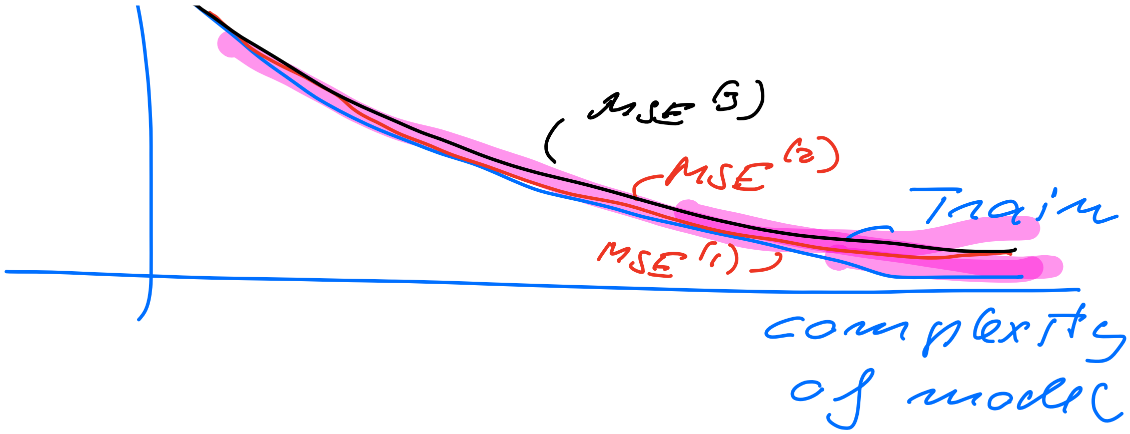
(iii) Repeat (i)-(ii) for a given number of Bootstraps M

(iv) compute for example MSE or other quantities

$$MSE = \frac{1}{M} \sum_{j=0}^{M-1} MSE_j$$

MSE





Cross-validation

- k-Folds (k-blocks)

$k = 5$ $\quad D = \{x_0, \dots x_{n-1}\}$

(i)

$x_0\ x_1 \dots$				x_{n-1}

Train Train Test$^{(1)}$

MSE$^{(1)}$

(ii)

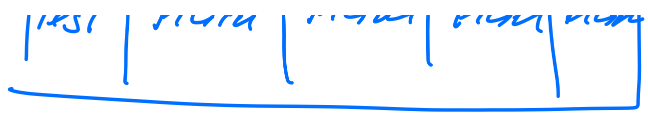
TRAIN	TRAIN	TEST	TRAIN	TRAIN

Test$^{(2)}$ $\Rightarrow$

MSE$^{(2)}$

(iii)

Test	Train	Train	Train	Train



CV

$$MSE = \frac{1}{K}\sum_{i=0}^{K-1} MSE_i$$

$K = 5 \sim 10$

LOOCV = Leave one out CV

$n = 100 \Rightarrow K = 100$, only one point as data set. $n = 10$

x_0	x_1	x_2	x_3	x_4	. . .	x_9

$\underbrace{x_0 - x_8}$ - training   test

$x_0, x_1, \dots x_7, x_9$ leave out x_8

$$X = \{x_0, x_1\}$$

$$y = \{y_0, y_1\}$$

$$z = f(x, y) = \text{Franke function}$$

Model : $f(x,y) \simeq \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 x \cdot y + \beta_5 y^2$

$$\begin{bmatrix} 1 & x_0 & y_0 \\ 1 & x_0 & y_1 \\ 1 & x_1 & y_0 \\ 1 & x_1 & y_1 \end{bmatrix}$$

x_0^2
x_1^2

$x_0 y_0$
$x_0 y_1$
$x_1 y_0$
$x_1 y_1$

y_0^2
y_1^2

TRAIN	VALIDATE	TEST