ML: Statistical learning

TMA4268
14.01.2019

i) problem types:

continuous

regression: quantitative output ← stock prices

classification: qualitative output ← disease or not

discrete

ii) aims: inference vs. prediction

interpret
understand

iii) set-up

supervised vs unsupervised learning

response variable available

regression classification

detect unknown patterns in data

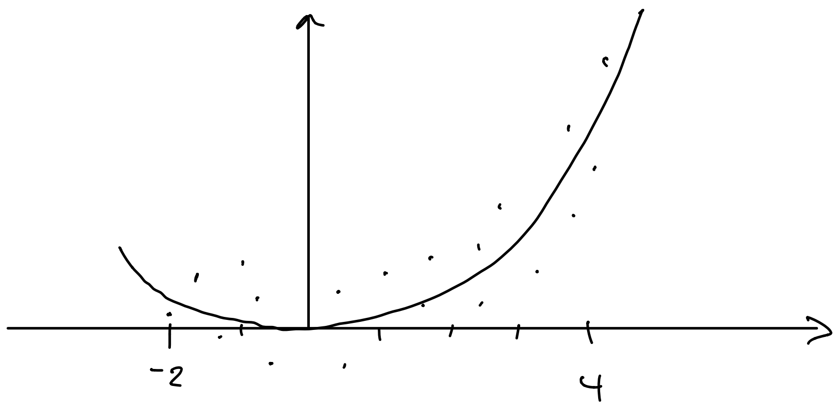
iv) type of method

parametric vs non-parametric

easy to interpret

Polynomial example: $Y = f(x) + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$

we here truth $f(x) = x^2$



$n = 61$

poly 1: $\overbrace{\beta_0 + \beta_1 x}^{f(x)}$: correct only at few positions

poly 2: $\beta_0 + \beta_1 x + \beta_2 x^2$: good

poly 10: $\beta_0 + \beta_1 x + \dots + \beta_{10} x^{10}$

poly 20: $\beta_0 + \dots + \beta_{20} \cdot x^{20}$

DATA

$Y = f(x) + \varepsilon$

estimate f

Training set — independent — Test set

$\hat{f}$ is found using the training set (not test)

Training set

mean squared error

MSE train

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$$

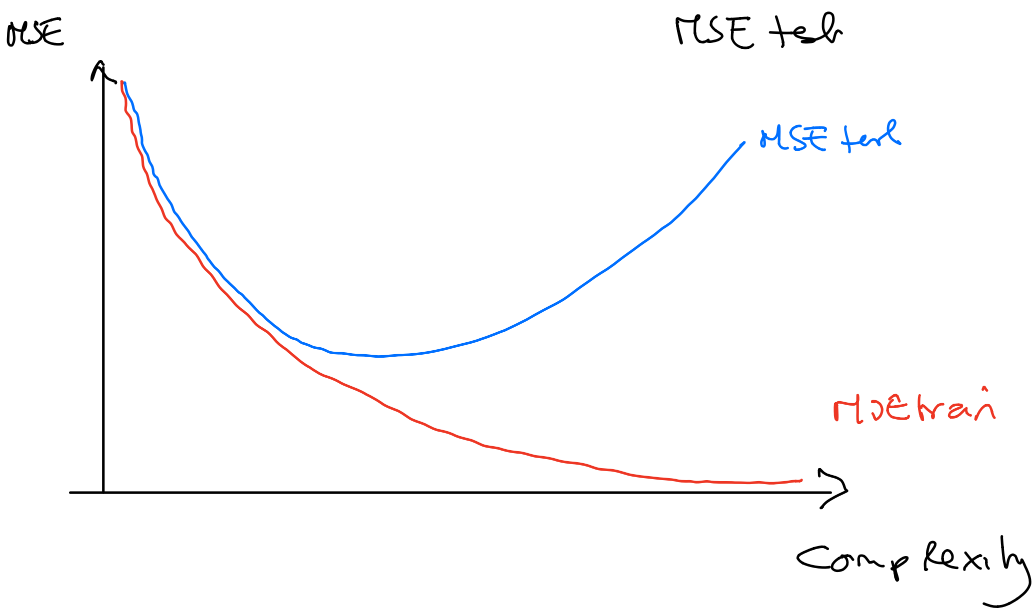
poly 20 has the lowest MSE train

Test set

MSE test ← n_0, x_{0j}, y_{0j}, $j = 1, \dots, n_0$

$$\frac{1}{n_0}\sum_{j=1}^{n_0}\left(y_{0j} - \hat{f}(x_{0j})\right)^2$$

poly 2 has the lowest MSE test



BIAS-VARIANCE TRADE-OFF

1) $Y = \underbrace{f(x)}_{\text{e.g polynom}} + \varepsilon$ where $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$

e.g polynom

2) Use training data to produce $\hat{f}$, for all possible x's

Y's

e.g the poly10 curve

$\hat{f}$ is a function of Ys

3) Focus on x_0 and want to look at Y at x_0

new value

4) "Expected test MSE at x_0"

RV = random variable

wrt Y

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

$E(Y + Y') = E(Y) + E(Y')$

$E(aY + b) = aE(Y) + b$

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

$-2 \cdot E(Y) \cdot E(\hat{f}(x_0))$

Y and $\hat{f}(x_0)$ are independent

new ... train

$$= \underbrace{Var(Y)}_{= Var(\varepsilon)} + \underbrace{E(Y)^2}_{f(x_0)^2} - 2 \cdot \overbrace{E(Y)}^{f(x_0)} \cdot E(\hat{f}(x_0)) + Var(\hat{f}(x_0)) + E(\hat{f}(x_0))^2$$

$Var(Y) = E((Y - E(Y))^2)$

$= E(Y^2) - (E(Y))^2$

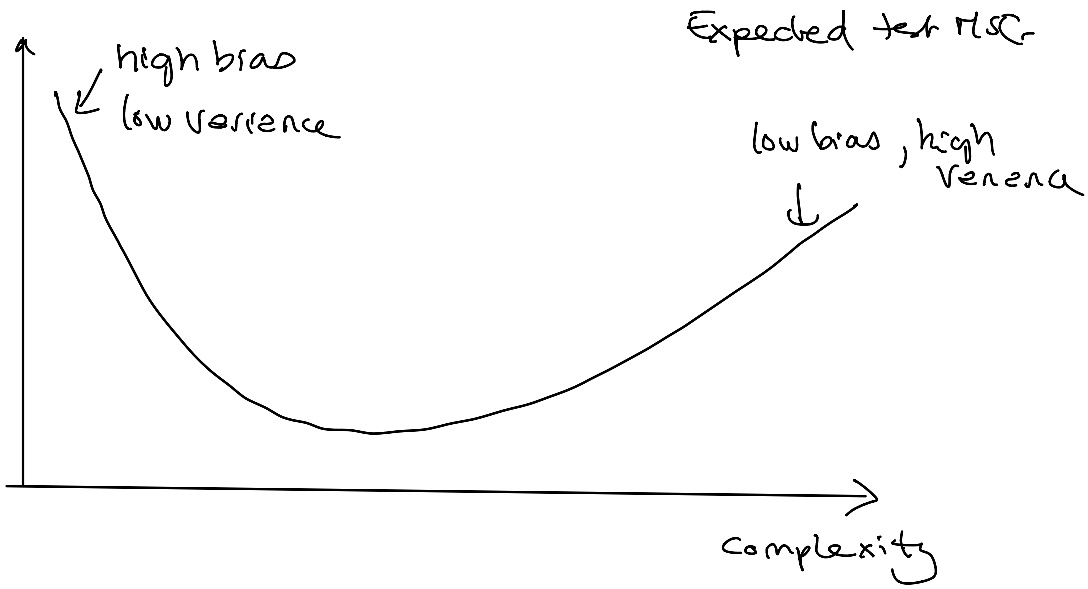
$E(Y^2) = Var(Y) + E(Y)^2$

$$= \overbrace{Var(\varepsilon)}^{\text{irreducible error}} + \overbrace{Var(\hat{f}(x_0))}^{\text{variance}} + \overbrace{[E(\hat{f}(x_0)) - f(x_0)]^2}^{\text{bias}^2}$$

$Y = f(x_0) + \varepsilon \Rightarrow E(Y) = f(x_0)$

$E(\varepsilon) = 0$

4


Expected test MSE
high bias
low variance
low bias, high variance
complexity