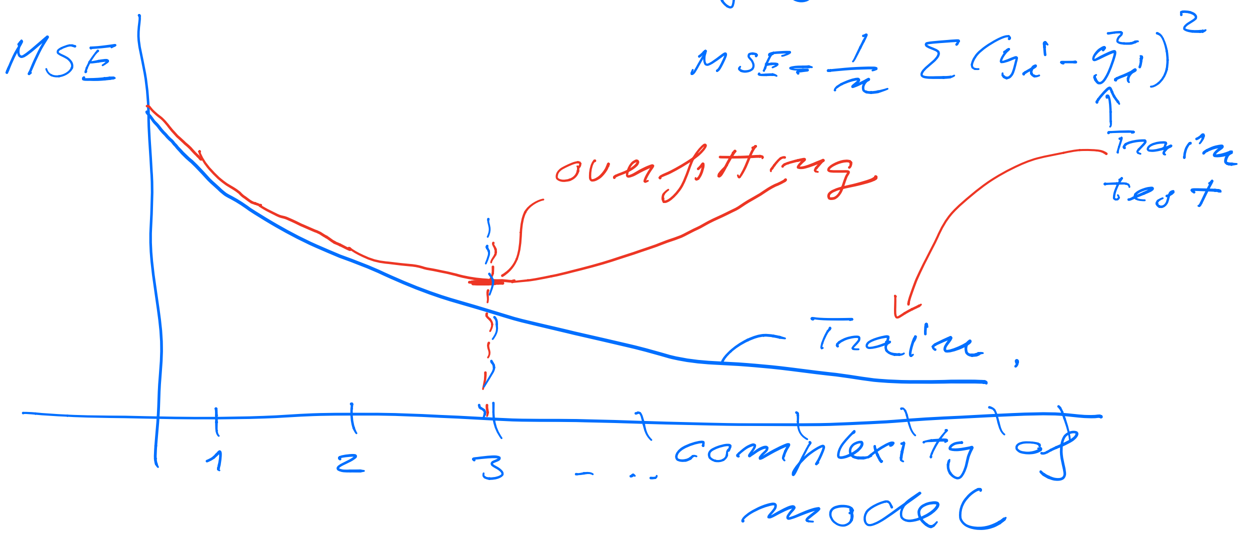
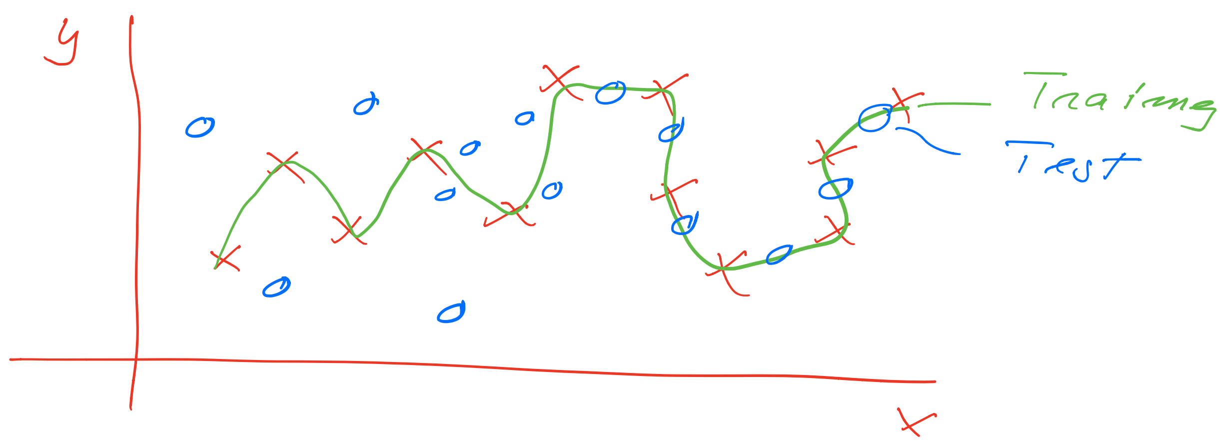
Some Elements from statistics

- split in training and test
- Have a polynomial approx

$$\tilde{y}_i = \tilde{y}(x_i) = \sum_{j=0}^{p-1} \beta_j x_i^j$$

$$MSE = \frac{1}{n} \sum (y_i - \tilde{y}_i)^2$$





Moments

PDF = p(x)

continuous case

$$\mathbb{E}[x^n] = \int x^n p(x)\,dx$$

mean value $\mu = \mathbb{E}[x]$

Discrete pdf.

$$\mu = \mathbb{E}[x] = \sum_{i=1}^{n} p(x_i)\, x_i$$

$$\mathbb{E}[x^n] = \sum_{i=1}^{n} p(x_i)\, x_i^n$$

variance

$$\sigma^2 = var[x] = \int (x-\mu)^2 p(x)\,dx$$

$$= \mathbb{E}[x^2] - \mu^2$$

$$var[x] = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

often $p(x)$ is unknown.

$\Rightarrow$ sample mean

$$\mu \neq \bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

sample variance

$$\sigma^2 \neq \bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{\mu})^2$$

covariance (sample)

$$\sigma(x,y) = \boxed{\frac{1}{n} \sum_{i=0}^{n} (x_i - \bar{\mu}_x)(y_i - \bar{\mu}_y)}$$

$\sigma(x,y)$ (exact $p(x,y)$)

$$= \iint dx\,dy\; p(x,y)\,(x-\mu_x)(y-\mu_y)$$

$$= \mathbb{E}[xy] - \mathbb{E}[x]\,\mathbb{E}[y]$$

$$= \iint dx\,dy\; p(x,y)\, x \cdot y$$

$$- \mu_x \cdot \mu_y$$

if x and y are iid

$$p(x,y) = p(x)\,p(y)$$

$$\iint dx\,dy\; p(x)\,p(y)\, x \cdot y =$$

$$\int dx\, p(x)\, x \int dy\, p(y)\, y$$

$$= \mu_x \cdot \mu_y \Rightarrow$$

covariance $= \sigma(x,y) = 0$

$$x = y \Rightarrow \sigma(x,y) = \sigma(x,x)$$

$$= \sigma^2(x)$$

Three variables x, y, z

$$\mathbf{x} = [x_0, x_1, \ldots x_{n-1}]$$

$$\mathbf{y} = [y_0, y_1, \ldots \cancel{x}y_{n-1}]$$

$$\mathbf{z} = [z_0, z_1 \ldots z_{n-1}]$$

covariance matrix

variance

$$
\begin{bmatrix}
\boxed{\sigma(xx)} & \sigma(xy) & \sigma(xz) \\
\sigma(xy) & \sigma(yy) & \sigma(yz) \\
\sigma(xz) & \sigma(yz) & \sigma(zz)
\end{bmatrix}
$$

correlation matrix = covariance divided by the variance ?

$$
\begin{bmatrix}
1 & x & x \\
x & 1 & x \\
x & x & 1
\end{bmatrix}
$$