1.

$$f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \le \sum_{i=1}^{M}\lambda_i f(x_i)$$

<pf>

① for $M=1$ $\quad f(\lambda a+(1-\lambda)b) \le \lambda f(a)+(1-\lambda)f(b)$ 成立

② 假設 $M=M$ 時成立：

$$f\left(\sum_{i=1}^{M}\lambda_i x_i\right) \le \sum_{i=1}^{M}\lambda_i f(x_i)$$

③ $M=M+1$ 時，

$$\begin{aligned}
f\left(\sum_{i=1}^{M+1}\lambda_i x_i\right) &= f\left(\sum_{i=1}^{M}\lambda_i x_i + \lambda_{M+1}x_{M+1}\right) \\
&= f\left(\lambda_{M+1}x_{M+1} + (1-\lambda_{M+1})\sum_{i=1}^{M}\frac{\lambda_i}{(1-\lambda_{M+1})}x_i\right) \\
&\le \lambda_{M+1}f(x_{M+1}) + (1-\lambda_{M+1})\, f\left(\sum_{i=1}^{M}\frac{\lambda_i}{(1-\lambda_{M+1})}x_i\right)
\end{aligned}$$

④ $\sum_{i=1}^{M+1}\lambda_i = 1 \quad \therefore \sum_{i=1}^{M}\frac{\lambda_i}{(1-\lambda_{M+1})} = \frac{1}{(1-\lambda_{M+1})}(1-\lambda_{M+1}) = 1$

$$\begin{aligned}
\Rightarrow f\left(\sum_{i=1}^{M+1}\lambda_i x_i\right) &\le \lambda_{M+1}f(x_{M+1}) + (1-\lambda_{M+1})\, f\left(\sum_{i=1}^{M}\frac{\lambda_i}{(1-\lambda_{M+1})}x_i\right) \\
&\le \lambda_{M+1}f(x_{M+1}) + (1-\lambda_{M+1})\sum_{i=1}^{M}\lambda_i f(x_i) \\
&= \sum_{i=1}^{M+1}\lambda_i f(x_i)
\end{aligned}$$

得証

SEASON

2.

NO.

DATE

$$H[x] = \frac{1}{2}\{1+\ln(2\pi\sigma^2)\}$$

<p.f> $\quad H[x] = -\int p(x)\ln p(x)\,dx$

$$\begin{aligned}
&= -\int \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}\cdot\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= \frac{1}{2}\left(\ln(2\pi\sigma^2) + \frac{1}{\sigma^2}\int \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}(x-\mu)^2\,dx\right) \\
&= \frac{1}{2}\left(\ln(2\pi\sigma^2) + \frac{1}{\sigma^2}\cdot\sigma^2\right) \\
&= \frac{1}{2}\left(\ln(2\pi\sigma^2) + 1\right)
\end{aligned}$$

3.

$$KL(p\|q) = -\int p(x)\ln q(x)\,dx - \left(-\int p(x)\ln p(x)\,dx\right)$$

$$= -\int p(x)\ln\frac{q(x)}{p(x)}\,dx$$

$$= -\int p(x)\ln\frac{\frac{1}{\sqrt{2\pi s^2}}e^{-\frac{(x-m)^2}{2s^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}\,dx$$

$$= -\int p(x)\left\{\ln\left(\frac{\sigma}{s}\right) + \left(-\frac{(x-m)^2}{2s^2} + \frac{(x-\mu)^2}{2\sigma^2}\right)\right\}dx$$

$$= -\left\{\ln\left(\frac{\sigma}{s}\right) - \frac{1}{2s^2}\int (x-m)^2 p(x)\,dx + \frac{1}{2\sigma^2}\int (x-\mu)^2 p(x)\,dx\right\}$$

$$= -\ln\left(\frac{\sigma}{s}\right) - \frac{1}{2} + \frac{1}{2s^2}E[(x-m)^2] \quad 其中\ E[(x-m)^2] = \sigma^2+\mu^2-2m\mu+m^2$$

$$= -\ln\left(\frac{\sigma}{s}\right) - \frac{1}{2} + \frac{1}{2s^2}\left(\sigma^2+\mu^2-2m\mu+m^2\right) \#$$

4.

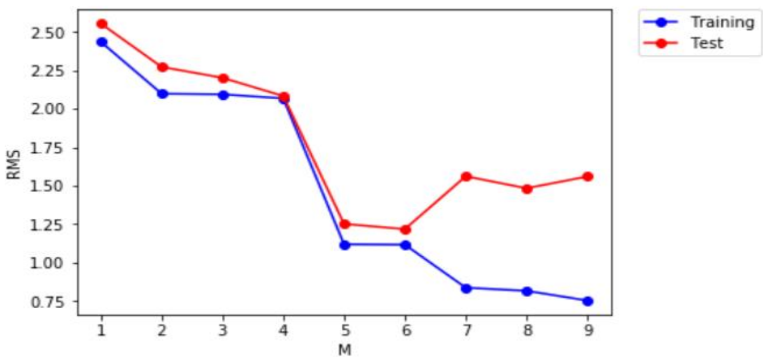
(1)

For training data

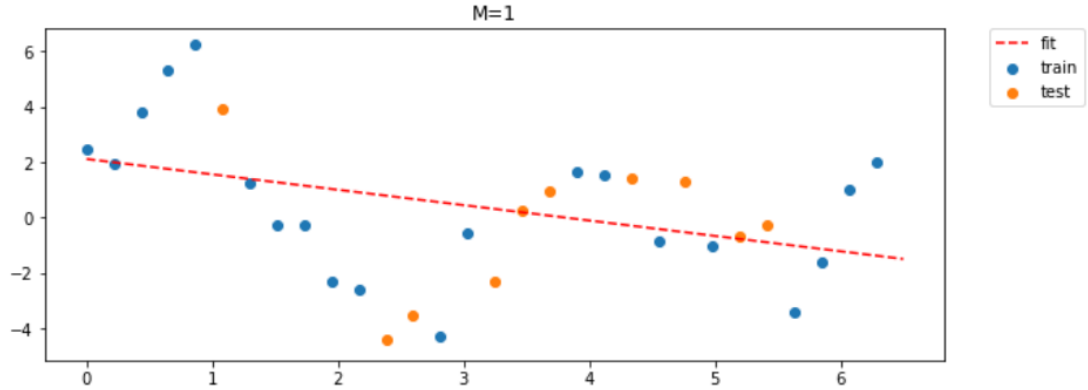
	M	RMSE
0	1	2.436250
1	2	2.098471
2	3	2.094265
3	4	2.067801
4	5	1.119144
5	6	1.116632
6	7	0.836488
7	8	0.816068
8	9	0.752874

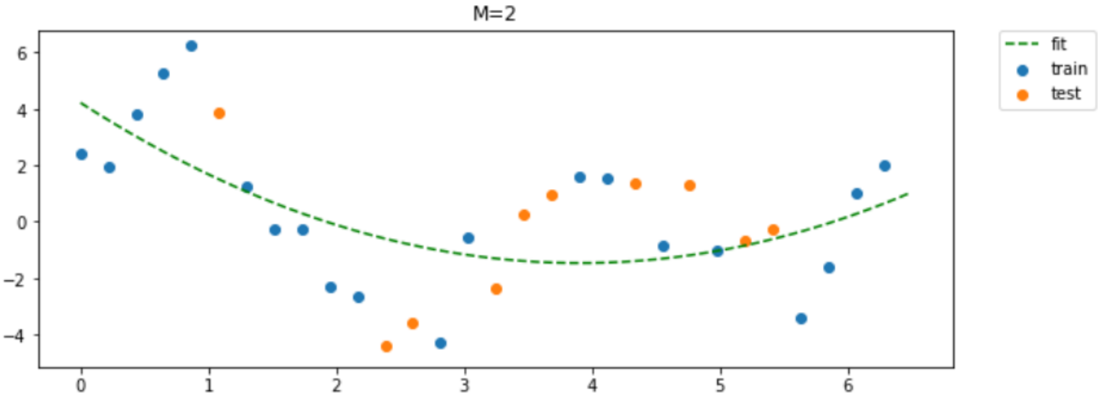
For testing data

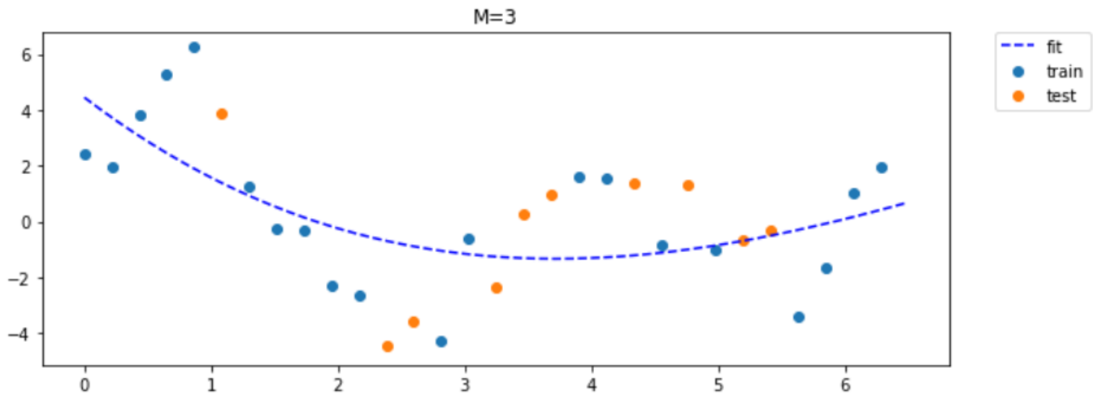
	M	RMSE
0	1	2.555124
1	2	2.272434
2	3	2.201998
3	4	2.083373
4	5	1.251469
5	6	1.217202
6	7	1.561197
7	8	1.483158
8	9	1.559494

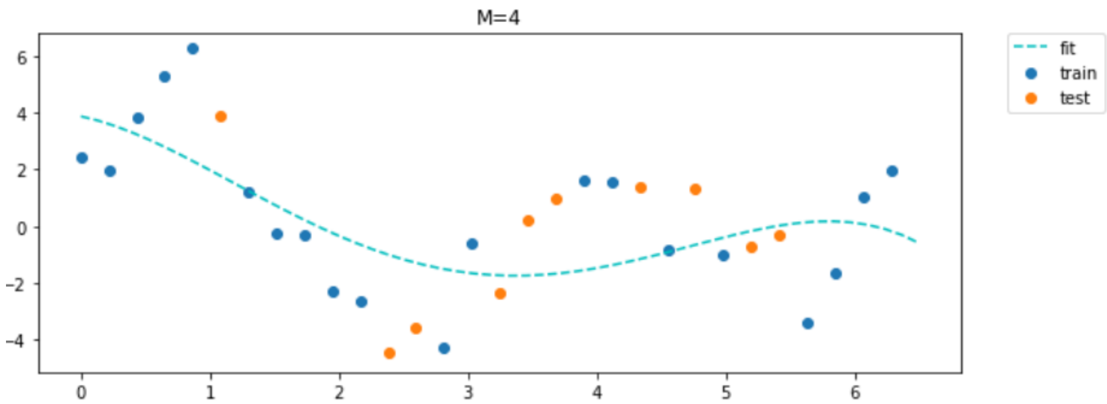


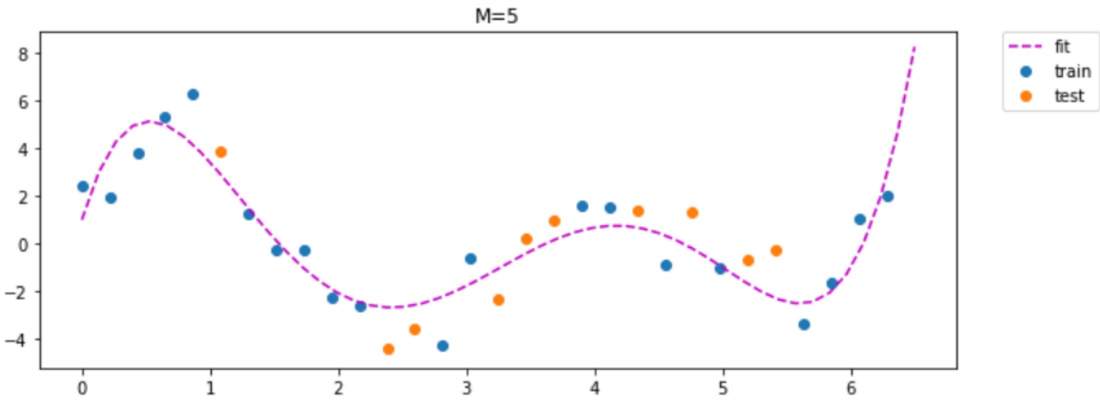
(2)

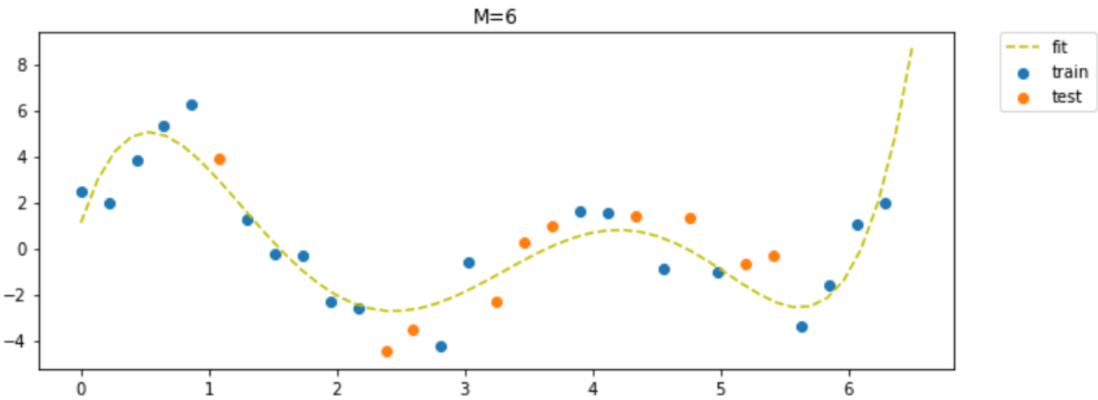

M=1
fit
train
test
6
4
2
0
−2
−4
0 1 2 3 4 5 6

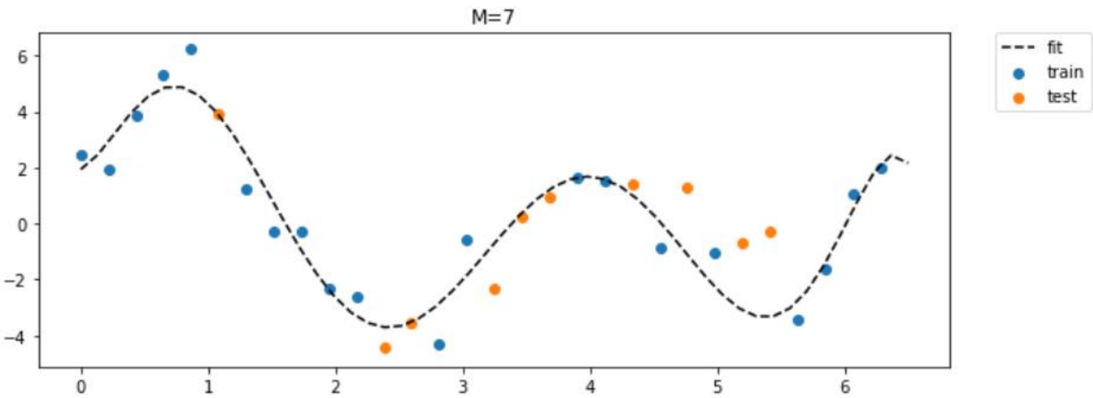

M=2
fit
train
test
6
4
2
0
−2
−4
0 1 2 3 4 5 6

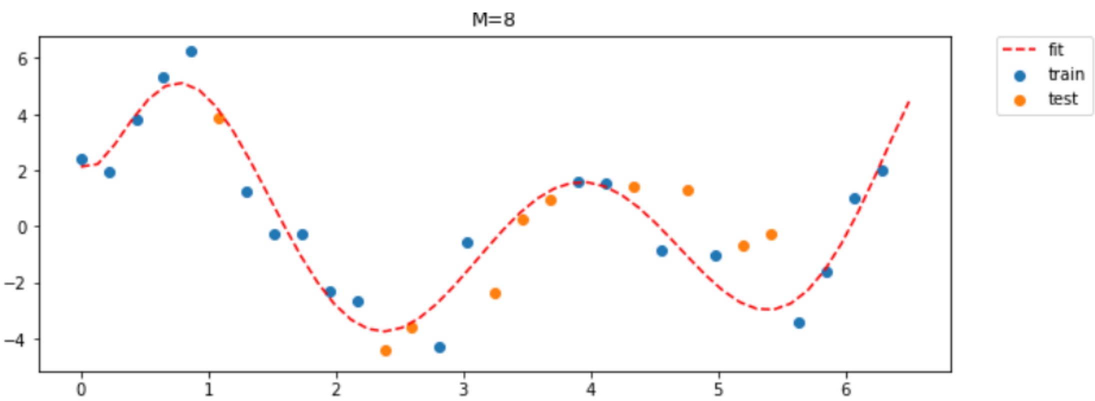

M=3
fit
train
test
6
4
2
0
−2
−4
0 1 2 3 4 5 6

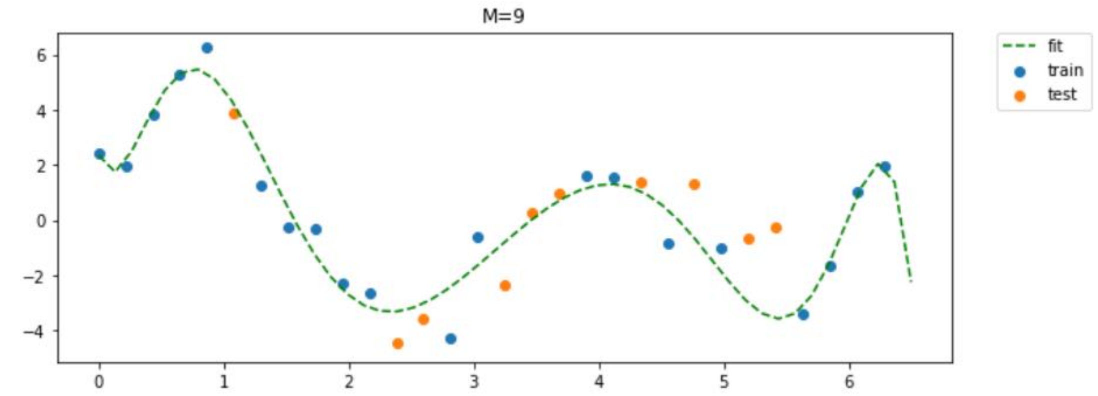

M=4
fit
train
test
6
4
2
0
−2
−4
0 1 2 3 4 5 6


M=5
fit
train
test
8
6
4
2
0
−2
−4
0
1
2
3
4
5
6

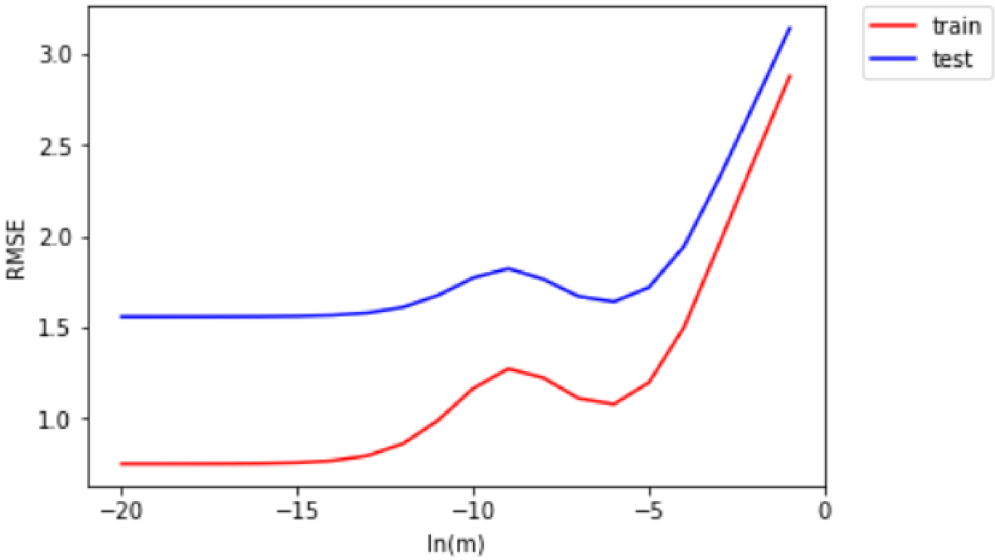

M=6
fit
train
test
8
6
4
2
0
−2
−4
0
1
2
3
4
5
6


M=7
fit
train
test
6
4
2
0
−2
−4
0
1
2
3
4
5
6


M=8
fit
train
test
6
4
2
0
−2
−4
0
1
2
3
4
5
6



(3)



5.

(1)

For M=1

rmse_trainnig = 1.50234684543

rmse_testing = 1.49595335281

For M=2

rmse_trainnig = 2.02982321837e+23

rmse_testing = 2.02982321833e+23

(2)

For attribute1 : rms = 12.074599802119355

For attribute2 : rms = 1.7340422684523498

For attribute3 : rms = 1.621134279263593

For attribute4 : rms = 2.0865167589303875

The attribute3 has the smallest rms error.