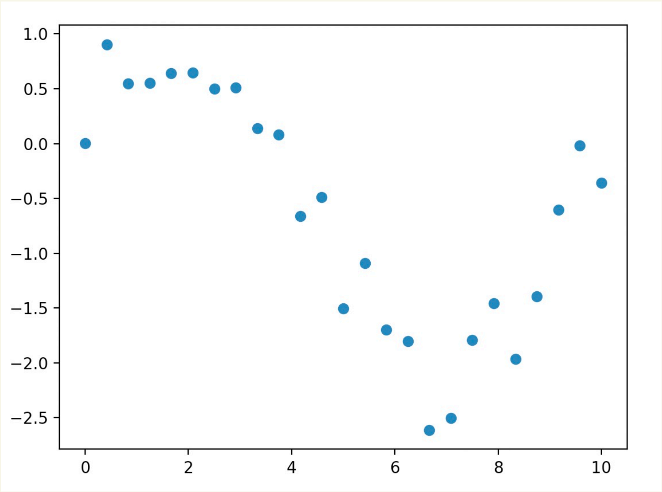
1.

(a) $J(w, D) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{f}(x_i))^2 = \frac{1}{N} \sum_{i}^{N} (y_i - \sum_{m=0}^{d} w_m x_i^m)^2$

(b) $J(\underline{w}, D) = \frac{1}{N} \| (\underline{y} - \underline{w}^T \underline{\phi}) \|_2^2$

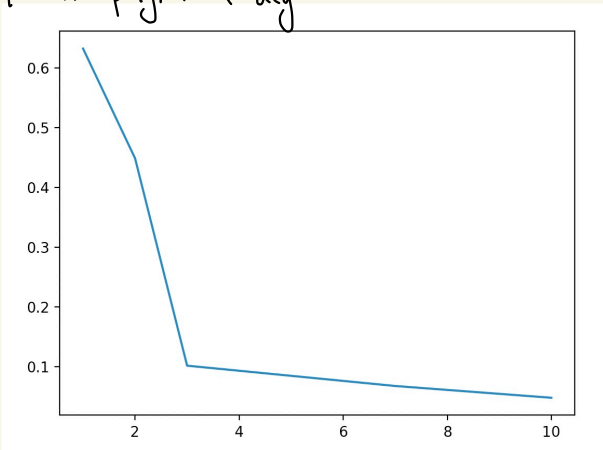
(c)



d.

```
w1 = [[ 0.55787413]
      [-0.23494215]]
w2 = [[ 1.40766343]
      [-0.76698414]
      [ 0.0532042 ]]
w3 = [[ 0.18301369]
      [ 0.87201349]
      [-0.36501611]
      [ 0.02788135]]
w7 = [[ 1.39650432e-01]
      [ 1.39453742e+00]
      [-1.22891445e+00]
      [ 5.21807626e-01]
      [-1.25503229e-01]
      [ 1.52807320e-02]
      [-8.47085275e-04]
      [ 1.60809536e-05]]
w10 = [[ 3.13911344e-03]
       [ 6.55154723e+00]
       [-1.64867156e+01]
       [ 1.80744823e+01]
       [-1.06467700e+01]
       [ 3.72266260e+00]
       [-8.10845713e-01]
       [ 1.11082547e-01]
       [-9.29962912e-03]
       [ 4.34434565e-04]
       [-8.67416690e-06]]
```

e. 10 th polynomial degress has the smallest MSE.

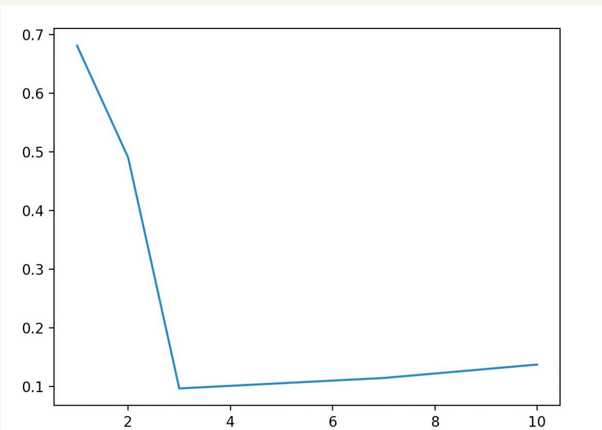


f.

MSE(W_1) MSE(W_2) MSE(W_3) MSE(W_7)

[0.68070979] [0.49014011] [0.09673217] [0.11453445]
[0.13738823]

MSE(W_{10})



3rd polynomial degree is the best

g. $\lambda = 10^{-5}$

```
[array([[ 1.39722954e-01],
       [ 1.39401406e+00],
       [-1.22822857e+00],
       [ 5.21447110e-01],
       [-1.25409398e-01],
       [ 1.52678852e-02],
       [-8.46198828e-04],
       [ 1.60566527e-05]]), array([[ 1.46572053e-01],
       [ 1.34468110e+00],
       [-1.16360339e+00],
       [ 4.87488371e-01],
       [-1.16572783e-01],
       [ 1.40582115e-02],
       [-7.62738908e-04],
       [ 1.37689100e-05]]), array([[ 2.67755109e-01],
       [ 5.05409144e-01],
       [-1.02239087e-01],
       [-5.65926056e-02],
       [ 2.24703565e-02],
       [-4.71641837e-03],
       [ 5.18727311e-04],
       [-2.10549261e-05]]), array([[ 2.83302412e-01],
       [ 2.51242694e-01],
       [ 6.83721533e-02],
       [-7.68300114e-02],
       [ 1.41197188e-02],
       [-2.14577212e-03],
       [ 2.64068828e-04],
       [-1.23729062e-05]]), array([[ 1.30648926e-01],
       [ 1.07289568e-01],
       [ 7.77265364e-02],
       [ 2.03327704e-02],
       [-3.10453463e-02],
       [ 6.03166173e-03],
       [-4.02835140e-04],
       [ 8.05983975e-06]])]
```

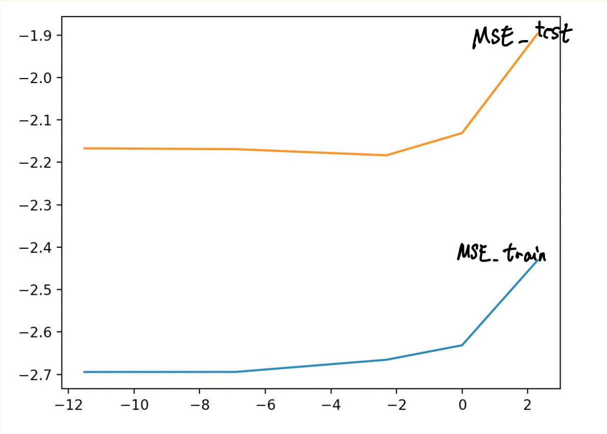
$\lambda = 10^{-3}$

$\lambda = 10^{-1}$

$\lambda = 1$

$\lambda = 10$

h.



MSE_test has minimal error when $\lambda = 10^{-1}$.

2. MLE error variance.

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i}^{N} (y_i - x_i^T \hat{w})^2$$

$$P(D \mid \sigma) = \sum_{i=1}^{N} P(y_i \mid x_i ; w_i)$$

$$NLL(\theta) = -\sum_{i=1}^{N} \log P(y_i \mid x_i, w_i)$$

$$= -\sum_{i=1}^{N} \log \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y_i - w^T x_i)^2}{2\sigma^2}} \right)$$

$$= -\left(-\frac{N}{2} \log 2\pi\sigma - \sum_{i=1}^{N} \frac{(y_i - w^T x_i)^2}{2\sigma^2} \right)$$

$$= \frac{N}{2} \log 2\pi\sigma - \frac{1}{2\sigma^2} RSS(w)$$

$RSS(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2$ is the sum of square error

$$MSE = E[RSS(w)]$$

$$= \frac{RSS(w)}{N}$$

3. (a)

$$\{ h_{odd}, h_5, h_{p5} \}$$

(b)

(i) likelihood $P(D|h_{p2}) = \left[\frac{1}{size(h)}\right]^N = \left(\frac{1}{6}\right)^1 = \frac{1}{6}$

posterior $P(h_{p2}|D) = \frac{P(D|h_{p2}) \cdot P(h_{p2})}{P(D)} = \frac{\frac{1}{6} \cdot \frac{3}{5}}{P(D)} = \frac{\frac{1}{10}}{P(D)}$

$= \frac{1}{10}$

(ii) likelihood $P(D|h_{p4}) = \frac{1}{3}$

posterior $P(h_{p4}|D) = \frac{P(D|h_{p4}) \cdot P(h_{p4})}{P(D)} = \frac{\frac{1}{3} \cdot \frac{2}{5}}{P(D)} = \frac{\frac{2}{15}}{P(D)} = \frac{2}{15}$

(iii) h_{p4} has a larger posterior

4. $P(\underline{w} \mid \underline{\underline{x}}, \underline{y}, \sigma^2)$

prior

$$= \boxed{\frac{P(\underline{y} \mid \underline{\underline{x}}, \underline{w}, \sigma^2) \cdot P(\underline{w} \mid \underline{\underline{x}}, \sigma^2)}{\int_w P(y \mid \underline{\underline{x}}, \underline{w}, \sigma^2) \cdot P(\underline{w} \mid \underline{\underline{x}}, \sigma^2)\, dw}}$$

likelihood

posterior