Problem 1)

The polynomial regression is given by the equation

$$f(x,\theta)=\theta_0+\theta_1 X+\theta_2 X^2+\ldots$$

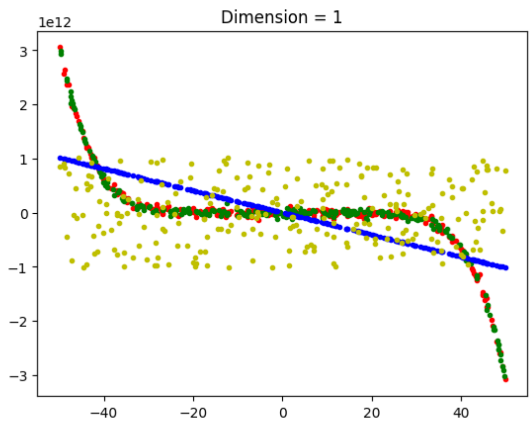
And the empirical risk is

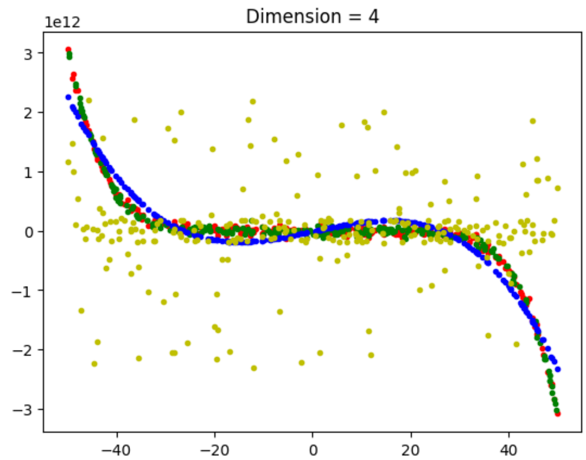
$$R(\theta)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left(y_i-f(x;\theta)\right)$$

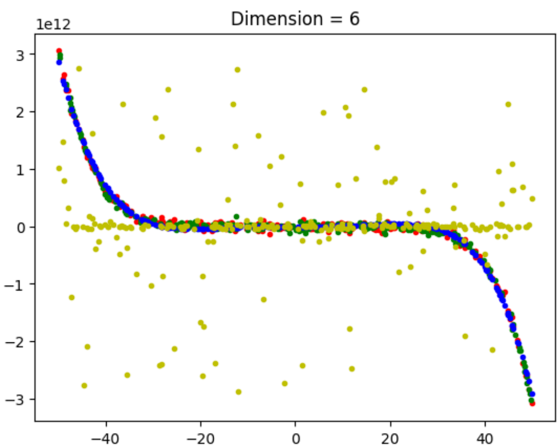
With partial derivative of empirical risk wtr θ,

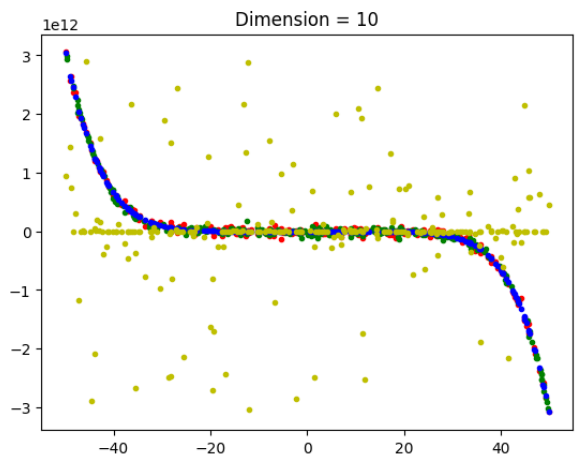
$\theta^{*} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}y$

By splitting the data into halves for test and train we get

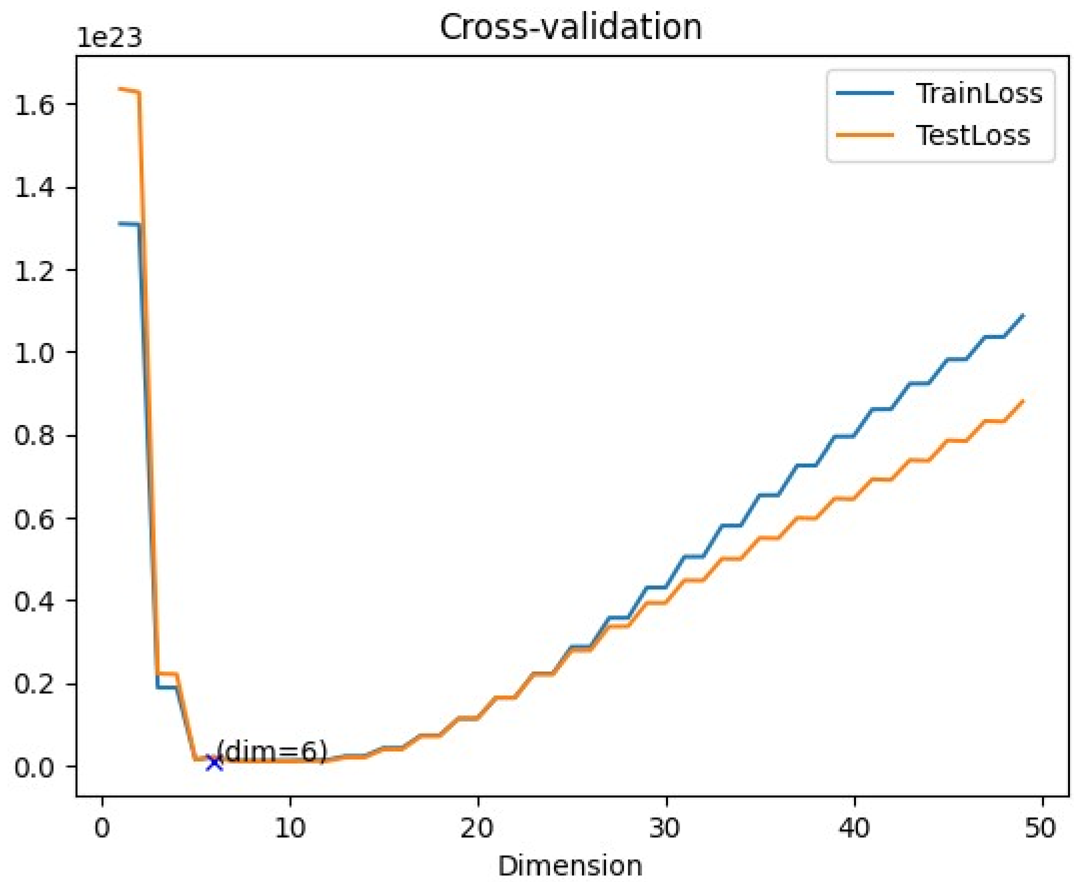








the cross validation graph is



Thus from the graph we can say that at dimension = 6 we get optimal θ^*

==

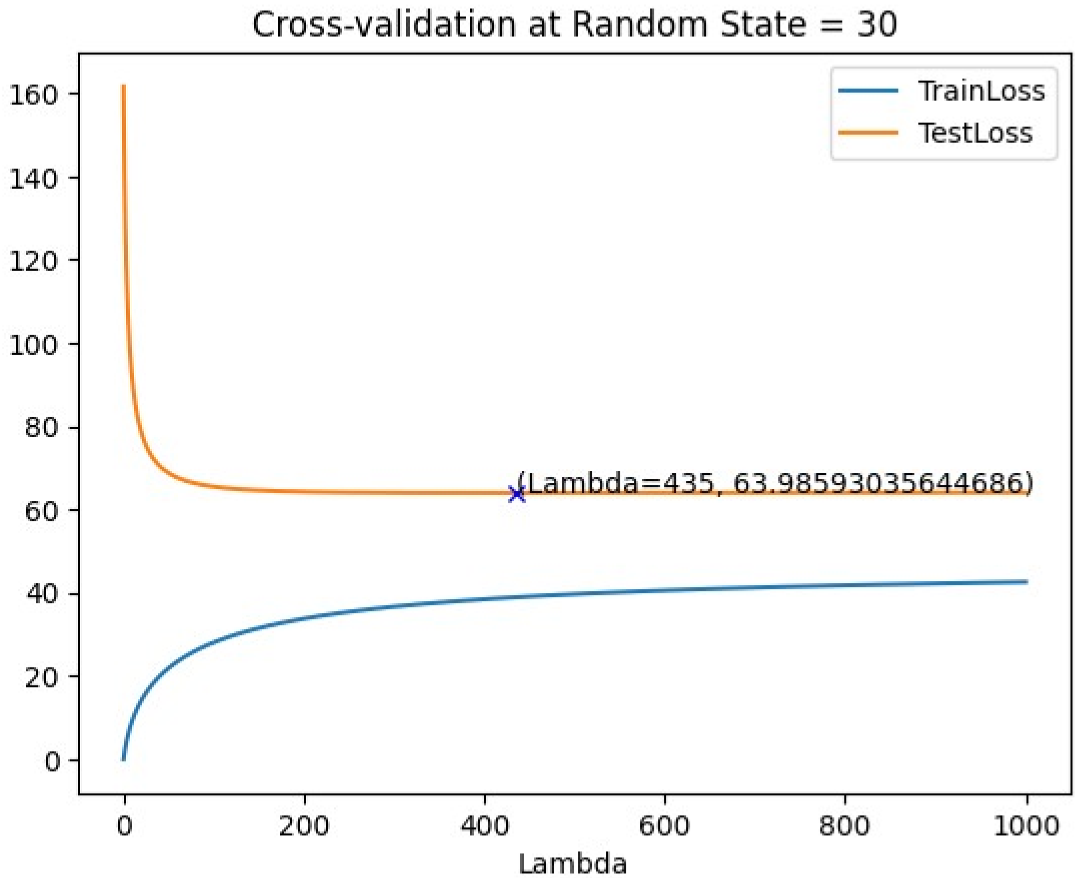
Problem 2:

Polynomial regression with l_2 Regularization

$R_{reg}(\theta) = R_{emp}(\theta) + \text{Penalty}(\theta)$

$$\frac{1}{N}\sum_{i=1}^{N} L\left(y_i - f\left(x;\theta\right)\right) + \frac{\lambda}{2N}$$

Where $\theta^* = (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T y$



At Lambda = 435, we get the least test loss.

==

Problem 3:
(I) proof of $g(-z)=1-g(z)$

$$g(z)=\frac{1}{\frac{1}{1+e^{-z}}}=\frac{e^z}{1+e^z}$$

$$g(-z)=\frac{1}{1+e^z}=\frac{1-e^z+e^z}{1+e^z}=\frac{1+e^z}{1+e^z}+\frac{e^z}{1+e^z}=1-\frac{e^z}{1+e^z}$$

$$g(-z)=1-g(z)$$

(II) Proof of $g^{-1}(y)=\ln\frac{y}{1-y}$

LHS = $g^{-1}(y)$

Assuming y =g(z)

LHS = $g^{-1}(g(\mathrm{z}))$ = z

RHS = $\ln(\frac{g(z)}{1-g(y)})=\ln\frac{\frac{e^z}{1+e^z}}{1-\frac{e^z}{1+e^z}}=\ln\frac{\frac{e^z}{1+e^z}}{\frac{e^z+1-e^z}{1+e^z}}$

= $\ln e^z = z$

RHS = z

LHS = RHS

==

Problem 8

The minimize the empirical risk of logistical regression:

$$R_{emp}(\theta)=\frac{1}{N}\sum_{i=1}^{N}(y_i-1)\log(1-f(x_i;\theta))-y_i\log(f(x_i;\theta))$$

$\nabla_\theta R(\theta)$ = $\frac{-1}{N}\sum_{i=1}^{N}(y_i\frac{1}{g(\theta^T x)}-(1-y_i)\frac{1}{1-g(\theta^T x)})\frac{\delta}{\delta\theta}g(\theta^T x)$

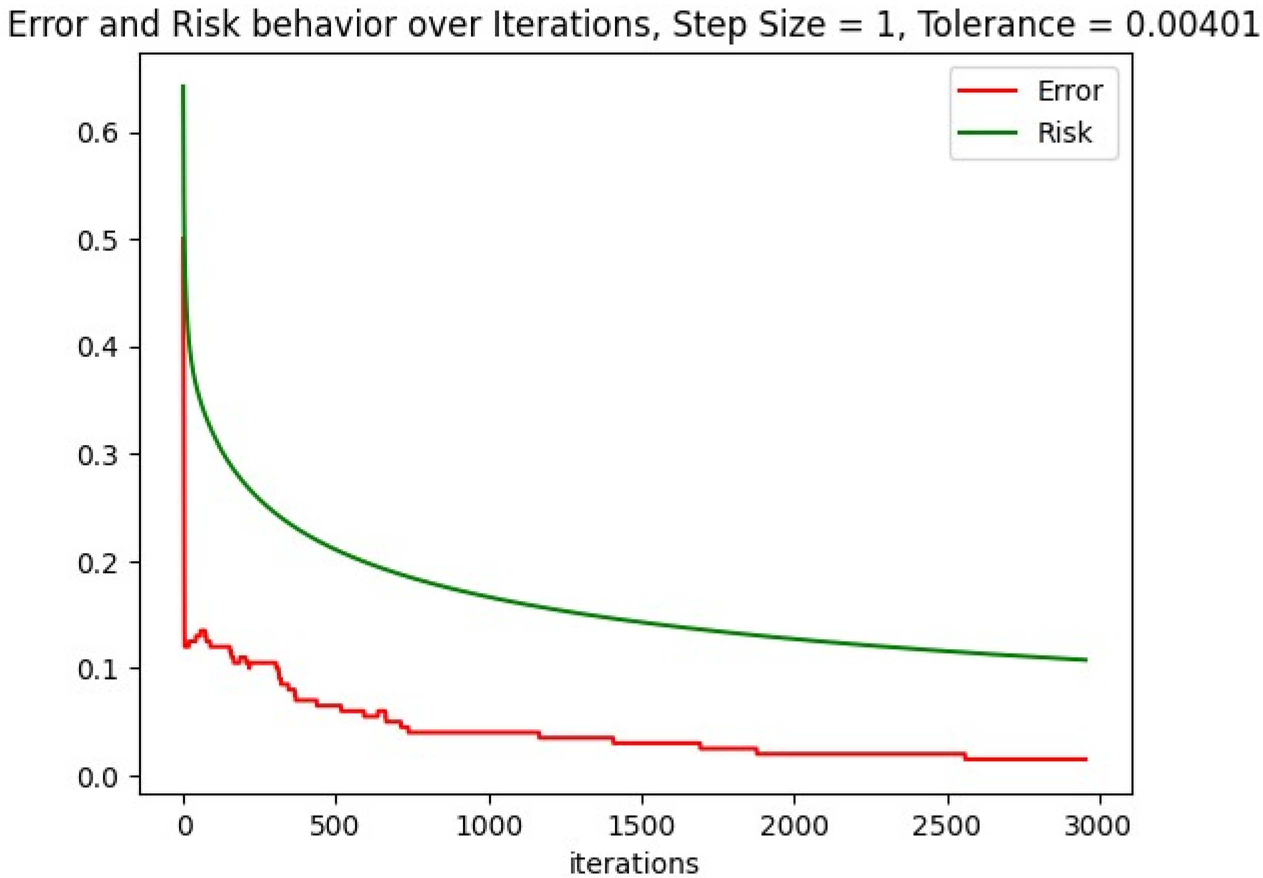
= $\frac{-1}{N}\sum_{i=1}^{N}(y_i\frac{1}{g(\theta^T x)}-(1-y_i)\frac{1}{1-g(\theta^T x)})g(\theta^T x)(1-g(\theta^T x))x$

= $\frac{-1}{N}\sum_{i=1}^{N}(y-g(\theta^T x))x$

To minimize the gradient

$\theta^{1+t} = \theta^t + \eta\nabla_\theta R(\theta^T)$

Iteration stops when the difference between the current and previous gradient is smaller then tolerance.



Minimum error where iteration stopped = 0.015
Minimum risk where iteration stopped = 0.10766697