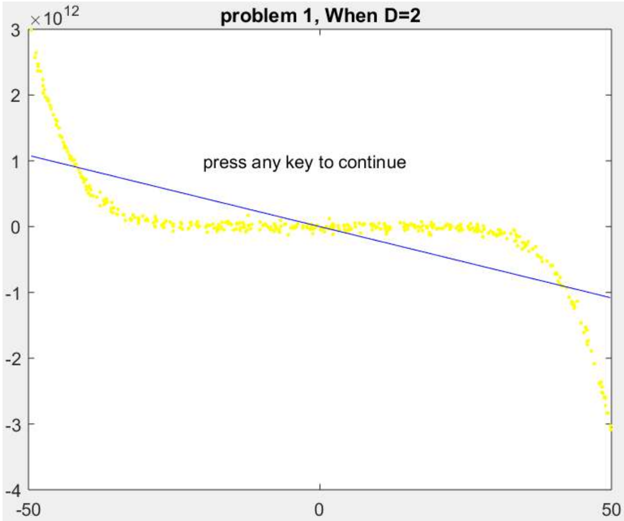
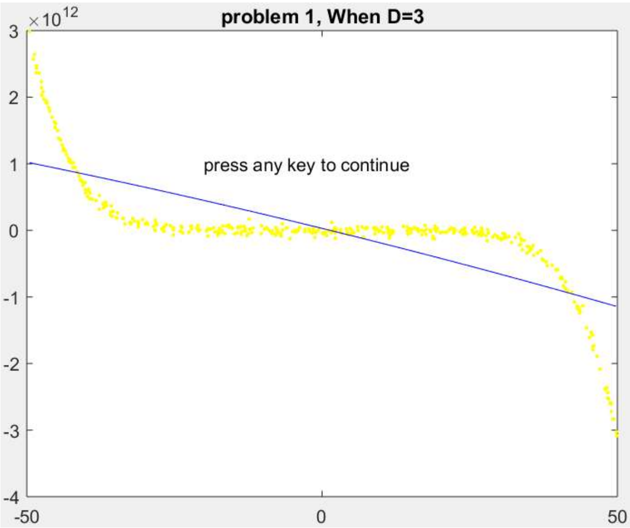
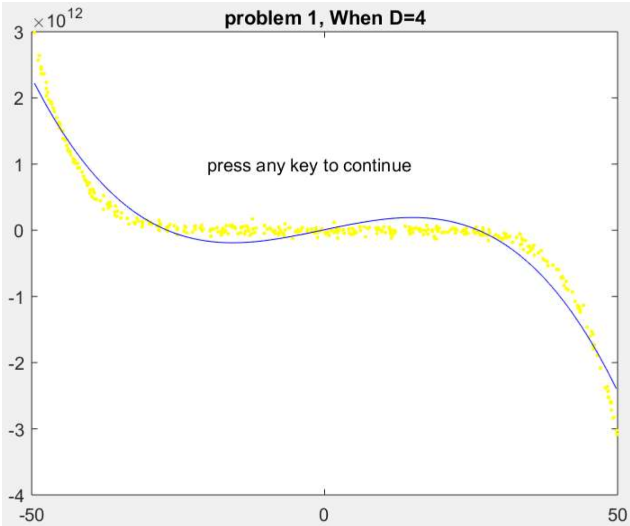
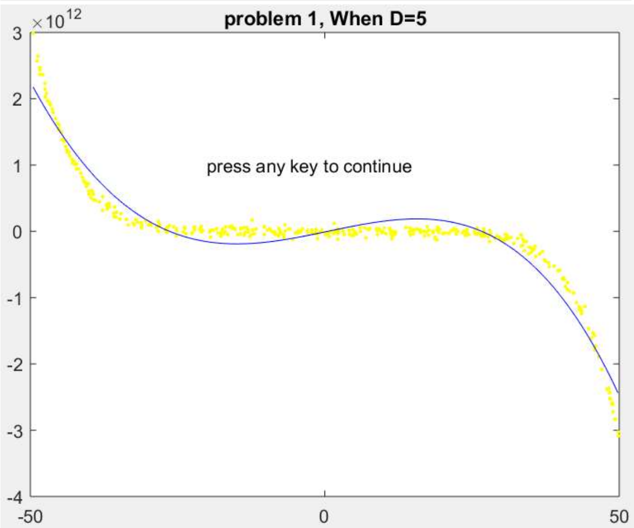
# Problem 1

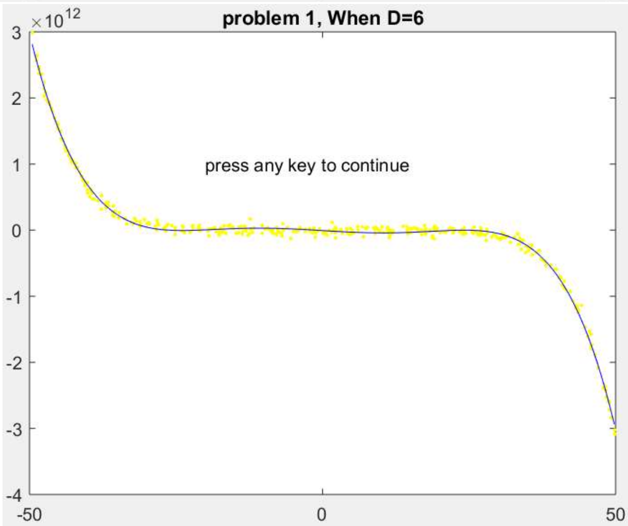
Polynomial models to the data for various choices for d (the degree of the polynomial):

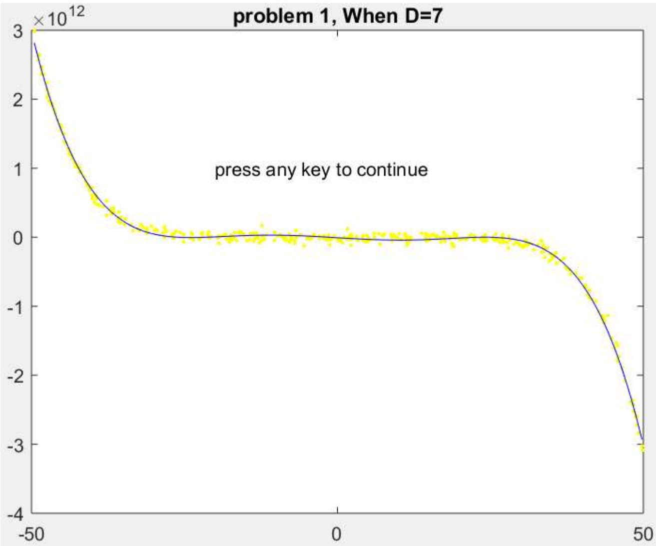



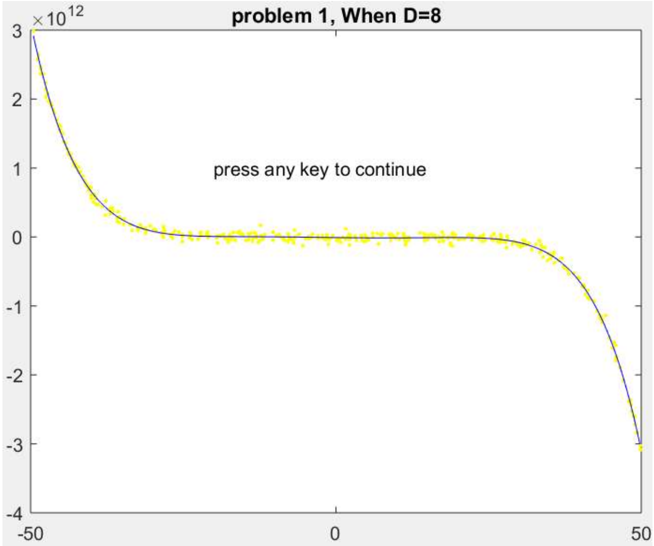



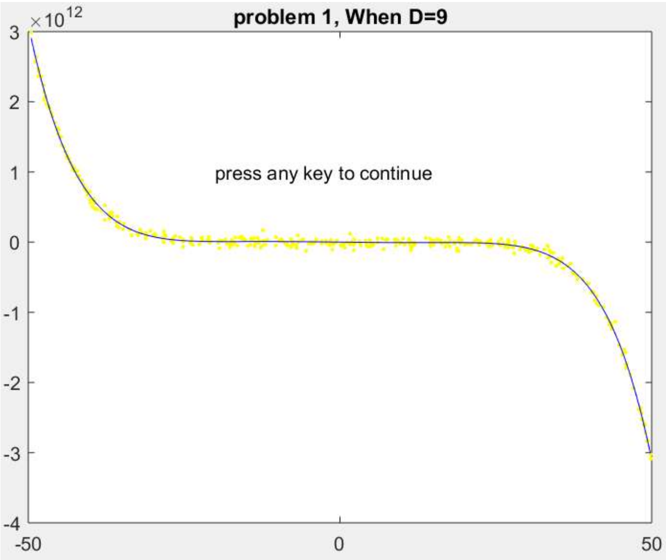



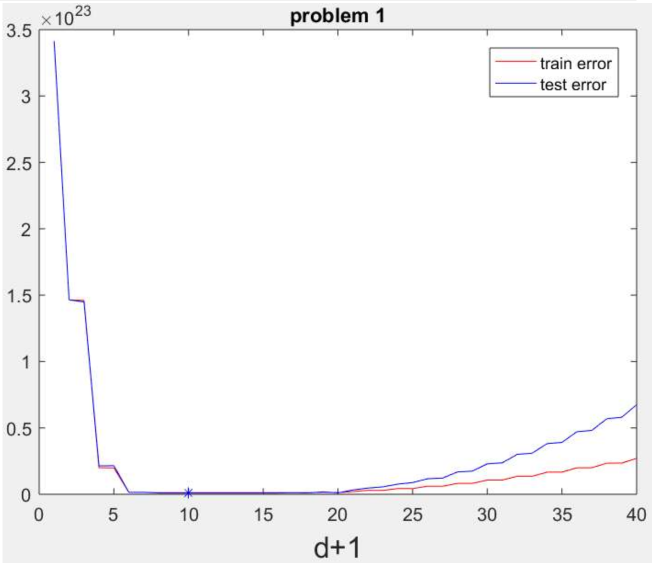






d = 0, θ = [ 9633660362.85 ]

d = 1, θ = [ -21662472992.07, -487530564.54 ]

d = 2, θ = [ -36210771.17, -21701599945.90, 29550498601.13 ]

d = 3, θ = [ -26397866.45, -28927403.04, 18603954039.65, 8429676489.32 ]

d = 5, θ = [ -34290.11, -26314771.72, 44891209.95, 18487304157.32, -9332361498.86 ]

d = 5, θ = [ -16217.47, -5453.28, 18647036.34, 9479887.64, -5253996631.56, -6731799659.82 ]

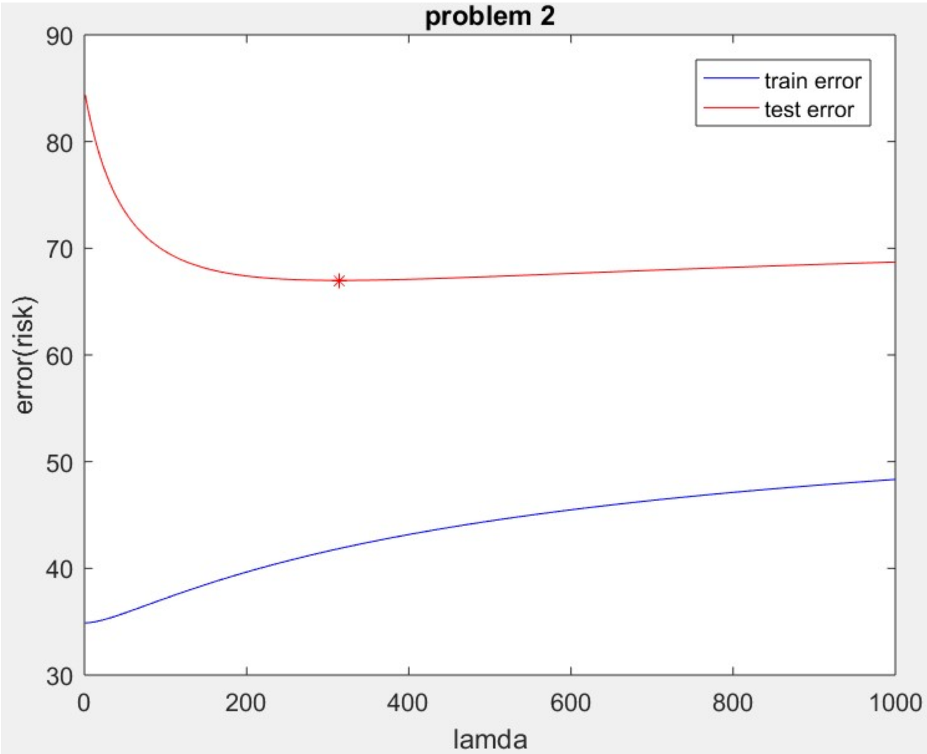
d = 6, θ = [ 2.11, -16220.10, -12625.73, 18651745.58, 15411662.10, -5257187644.35, -7415234103.16 ]

d = 7, θ = [ -3.44, 8.55, -2390.68, -31386.89, 3045117.05, 31309960.54, -1007329991.73, -8452246257.26 ]

d = 8, θ = [ -0.0053, -3.42, 25.27, -2465.73, -35309.19, 3114855.12, 11965757.61, -1013528741.59, 1606178.54 ]

The last figure is the result of the cross-validation. According to the computation by MATLAB, the best choice of d is 9. (d is the degree of the polynomial.)

## Problem 2




As can be seen in the plot above, increasing λ rapidly decreases the testing error (while increasing training error). The minimal λ value in this case was 314, though this is dependent on how the data set is split.

## Problem 3

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

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

② $\ln\frac{y}{1-y} = \ln y - \ln(1-y)$

$$= \ln\frac{1}{1+e^{-z}} - \ln\frac{1}{1+e^{z}}$$

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

Thus, $g^{-1}(y) = z = \ln\frac{y}{1-y}$.

## Problem 4

When use gradient descent, we need to know $\nabla_\theta R_{emp}$ first.

$$f = f(x;\theta) = \frac{1}{1+e^{-\theta^T x}}$$

$$R_{emp} = \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\log(1-f_i) - y_i\log f_i$$

then

$$\nabla_\theta R_{emp} = \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{d}{d\theta}\log\left(\frac{e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right) - y_i\frac{d}{d\theta}\log\left(\frac{1}{1+e^{-\theta^T x_i}}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{d}{d\theta}\left[-\log(1+e^{\theta^T x_i})\right] - y_i\frac{d}{d\theta}\left[-\log(1+e^{-\theta^T x_i})\right]$$

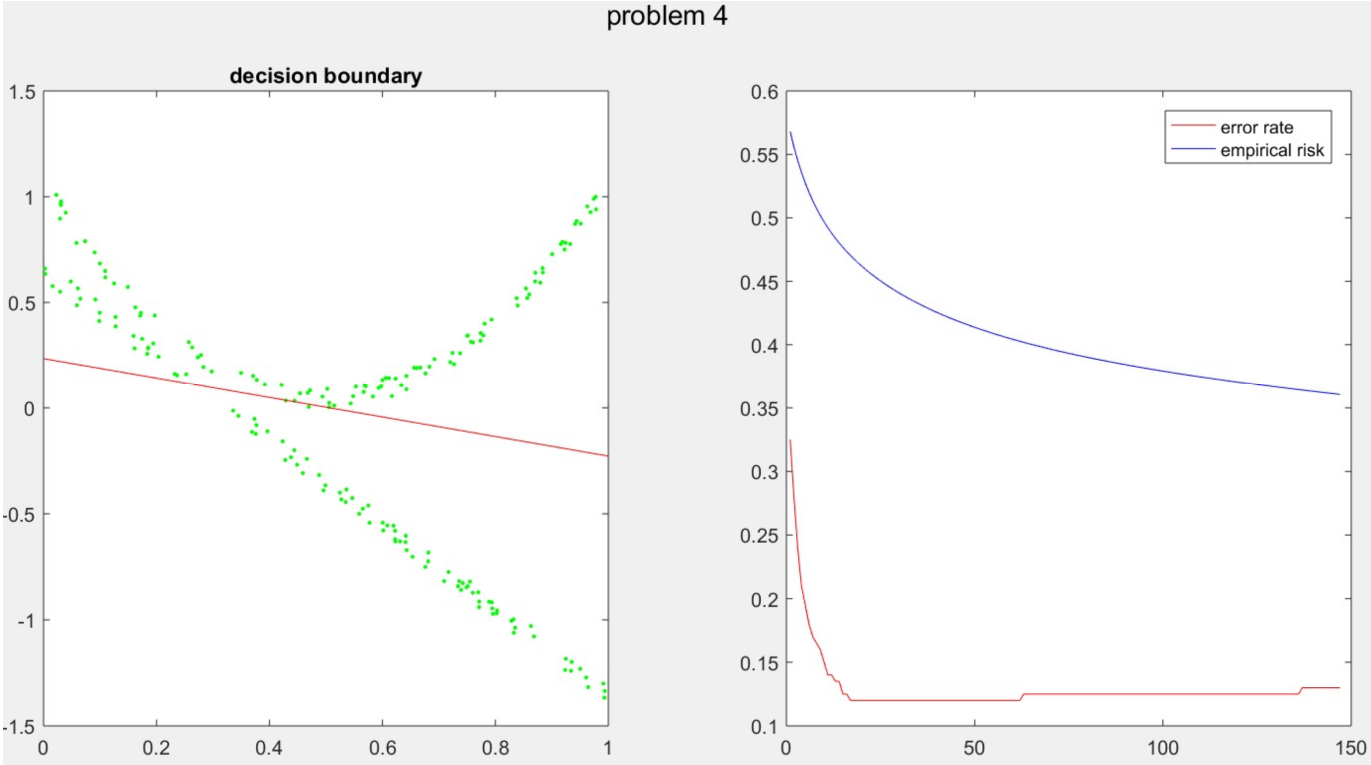
$$= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{-x_i e^{\theta^T x_i}}{1+e^{\theta^T x_i}} - y_i\frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}$$

$$= \frac{1}{N}\sum_{i=1}^{N}(y_i-1)\frac{-x_i}{1+e^{-\theta^T x_i}} - y_i\frac{x_i}{1+e^{\theta^T x_i}}$$

## Problem 4

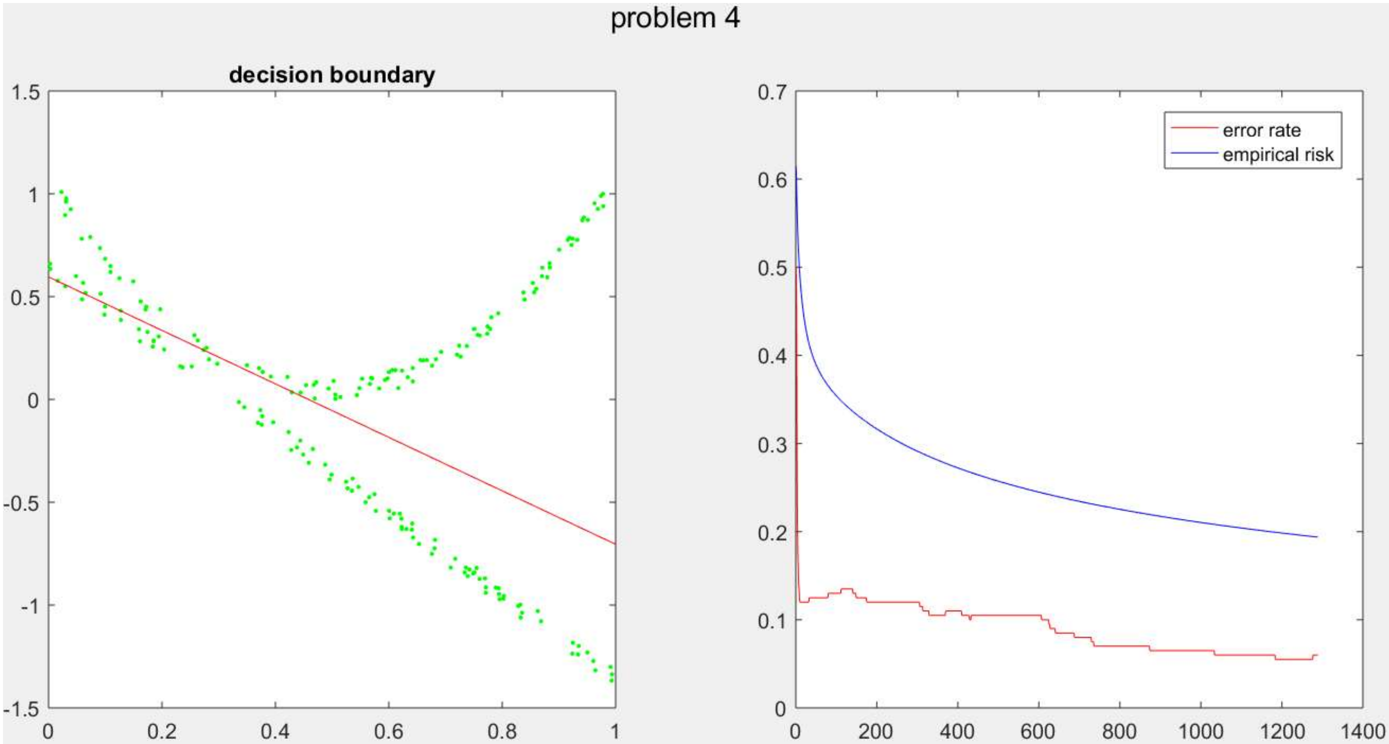
Step size η = 0.3, tolerance ε = 0.01

Linear boundary θ = [1.132, 3.217, -0.580]




Step size η = 0.5, tolerance ε = 0.005

Linear boundary θ = [9.295, 7.149, -4.258]



Step size η = 1, tolerance ε = 0.001

Linear boundary θ = [48.885, 26.897, -19.248]