Long Nguyen

1001705873

1)

degree = 1, lambda = 0

w0=40.2937

w1=-85.3182

w2=40.5272

w3=2.8325

w4=2934.2841

w5=-14575.7107

w6=2403.3571

w7=5.3809

w8=-1217.1594

w9=238.0055

w10=-8.3754

w11=-641.5481

w12=6.1993

w13=-395.2040

ID = 102, output = 25.3395, target value = 25.0000, squared error = 0.1153

degree = 2, lambda = 0

w0=166.3681

w1=-298.2493

w2=1754.4640

w3=-43.9698

w4=412.4657

w5=-58.9031

w6=2286.8364

w7=3101.1087

w8=4.3616

w9=-17152.5014

w10=389.2443

w11=-15204.6413

w12=970080.3976

w13=-14.4500

w14=100.9698

w15=-1787.9444

w16=65442.1147

w17=387.9337

w18=-4914.2598

w19=-23.3693

w20=15.3658

w21=-3571.2823

w22=62091.1718

w23=17.6688

w24=-22.6487

w25=-1020.4161

w26=12937.5971

ID = 102, output = 25.0664, target value = 25.0000, squared error = 0.0044

degree = 1, lambda = 1

w0=23.4505

w1=-4.9610

w2=20.0482

w3=-4.3727

w4=0.1951

w5=-0.0402

w6=2.1384

w7=-8.8236

w8=-0.3145

w9=1.3135

w10=-11.0302

w11=-2.7073

w12=12.2978

w13=-16.1952

ID =   102, output =   19.8046, target value =   25.0000, squared error = 26.9919

degree = 2, lambda = 1

w0=22.4499

w1=-4.7353

w2=-0.3711

w3=19.7559

w4=2.2267

w5=-4.2212

w6=-0.1121

w7=0.1956

w8=0.0003

w9=-0.0382

w10=-0.0001

w11=2.1226

w12=0.0387

w13=-8.4829

w14=-1.7052

w15=-0.3628

w16=-0.0073

w17=1.6790

w18=0.0491

w19=-6.8179

w20=-3.3553

w21=-2.6919

w22=-0.1457

w23=9.2713

w24=5.1126

w25=-16.0614

w26=-0.5957

ID = 102, output = 19.6992, target value = 25.0000, squared error = 28.0982

2) In the formula in slide 60, as λ grows so will $\frac{\lambda}{2}w^T w$. Therefore, to minimize $\tilde{E}_D(\mathbf{w})$, w would have to be approaching zero to prevent the regularization term from being dominant and causing the formula to be skewed and receiving the penalty of a large w. Thus, w = 0.

3) Using the formula: $E_D(w) = \frac{1}{2}\sum_{n=1}^{N}[t_n - f(x_n)]^2$

For $f(x) = 3.1x + 4.2$:

f(5.3) = 3.1(5.3) + 4.2 = 20.63

f(7.1) = 3.1(7.1) + 4.2 = 26.21

f(6.4) = 3.1(6.4) + 4.2 = 24.04

$[t_1 - f(x_1)]^2$ = (9.6 - 20.63)^2 = 121.66

$[t_2 - f(x_2)]^2$ = (4.2 - 26.21)^2 = 484.44

$[t_3 - f(x_3)]^2$ = (2.2 - 24.04)^2 = 476.9856

$E_D(w) = \frac{1}{2}(121.66 + 484.44 + 476.9856)$ = 541.54

For $f(x) = 2.4x - 1.5$:

f(5.3) = 2.4(5.3) - 1.5 = 11.22

f(7.1) = 2.4(7.1) - 1.5 = 15.54

f(6.4) = 2.4(6.4) - 1.5 = 13.86

$[t_1 - f(x_1)]^2$= (9.6 - 11.22)^2 = 2.62

$[t_2 - f(x_2)]^2$= (4.2 - 15.54)^2 = 128.60

$[t_3 - f(x_3)]^2$= (2.2 - 13.86)^2 = 135.96

$E_D(w) = \frac{1}{2}(2.62 + 128.60 + 135.96)$ = 133.59

The error for f(x) = 2.4x - 1.5 is much smaller than the other formula, therefore f(x) = 2.4x - 1.5 is the better solution.

4) Bob algorithm should not replace the standard algorithm, since calculating the most optimal lambda would make it so that the computation time will increase the larger the dataset is. Additionally, forcing lambda and w to be optimal may cause overfitting since it is determined by the training data alone and not the test data as well. This will cause the model to be able to generalize data.