

Hw3 (due on April/24 12:30PM)

Ques1 (40 points): Use dataset linedata01.txt and MatLab function (polyfit) to least-squares fit polynomials of degree 1, 2, 3, and 4 to data. Make plots that show observed and predicted data. Comment on your results and consider to address some questions listed below: Which line do you think best fits the observation? What criteria can you use to determine whether an overfitting occurs in your regression model?

Hints: Extend the linear regression model with additional terms

- a) $1+x$
- b) $1+x+x^2$
- c) $1+x+x^2+x^3$
- d) $1+x+x^2+x^3+x^4$

Plot a figure for each regression and point out the best regression model among ($1+x$, $1+x+x^2$, $1+x+x^2+x^3$, $1+x+x^2+x^3+x^4$).

Quest2 (20 points): Same as Q1 but using the data kernel method to redo the regression with

- a) $1+x+x^2$
- b) $1+x+e^x$

Comment on your results and consider whether replacing x^2 by e^x term helps the fitting.

Ques3 (40 points): Modify MatLab script, sample.m, to achieve a better fit to the Black Rock Forest temperature dataset (brf_temp.txt).

- 1) add one additional term x^2 in the regression model.
- 2) add additional periods of $T_y/2$, where T_y is the period of 1 year, in an attempt to better capture the shape of the (semi) annual variation
- 3) add additional periods of $T_y/2$ and $T_y/3$ together.

Do you think which one works best to achieve a better fit.