Homework policies: as in previous homework

Assignment

Exercise 1 (50 pts). Write a Matlab function `function [theta,b]=LinearRegression(X,Y)` that takes as inputs a $D \times n$ matrix $X$ whose columns represent $n$ points in $\mathbb{R}^D$, and a $n$ row vector $Y$ whose $i$-th entry represents the value of a function of interest evaluated at the $i$-th point. The outputs are two $D$-dimensional column vectors, $\theta$ and $b$, such that $\hat{Y} = \theta^T X + b$ is the least squares estimator from the $n$ input samples $(X,Y)$. Consider the following choices of $X$ and $Y$:

- $X$ a set of $n$ points in the $D$-dimensional unit cube, for all combinations of the values $n = 10, 100, 1000$ and $D = 1, 2, 10$, and choices of $Y(x) = \langle x, e_1 \rangle$, where $e_1$ is the first unit standard vector; $Y(x) = 10(\langle x, e_1 \rangle)^3 - \langle x, e_2 \rangle$; $Y(x) = ||x||_2$.

- $X$ is a set of $n$ handwritten digits 1 and $n$ handwritten digits 7, for $n = 10, 100, 1000$ (and $n = 5000$ if your computer can handle it) and $Y$ is the function equal to 1 on images of digit 1 and equal to 7 on images of digit 7.

- $X$ is a set of $n$ handwritten digits of each class 0, 1, ..., 9, for $n = 10, 100, 1000$, and $Y(x)$ is equal to the digit represented by the image $x$.

For each of the data sets above, construct a training data set $(X_{train},X_{test})$ with the specified sizes, and then a test set $(X_{test},Y_{test})$ of the same size. Use the training data to construct $\theta$ and $\hat{b}$. Then obtain $\hat{Y}_{train}$ and $\hat{Y}_{test}$ by applying the linear model to $X_{train}$ and $X_{test}$ respectively. Measure the mean square error on the training data and then on the test data, and plot it as a function of $n$ and $D$ where applicable. In the case of the digits, also compute the classification error, besides the mean square error, i.e. the number of wrongly predicted labels.

Finally, for each of the training sets above, add Gaussian noise $\mathcal{N}(0,(\max_i ||X(:,i)||)\sigma^2 D, I_D)$, with $\sigma = 0, 0.01, 0.1, 1$. Study how the regression results change with $\sigma$. 