

Due Date: Thursday, Oct 8, 2020.

Reading Assignment: Please read Chapters 5, 6, 8, 9, 13 from the textbook Shalev-Schwartz and Ben-David.

This is a computational assignment where we will study the problems of Underfitting vs. Overfitting, the usage of Test sets, and Regularization.

- Let $\mathcal{X} = [0, 1]$. Let $h_{TRUE}(x) := \cos 2\pi x$. Generate a Training Set of m samples as follows. First generate $\{x_i : 1 \leq i \leq t\}$, where each x_i is independently and uniformly distributed on $[0, 1]$. For each x_i , the noisy label is $y_i := h_{TRUE}(x_i) + w_i$, where the w_i are independently and $N(0, 0.01)$ distributed, i.e., normally distributed with mean 0 and variance 0.01. Independently also generate a separate Test set of t labeled samples in the same way.
 - For a Training Set of $m = 10$ samples, determine the least squares fit by polynomials $p_n(x) = w_0 + w_1x + w_2x^2 + \dots + w_nx^n$ of degrees less than or equal to $n = 0, 1, 2, \dots, 10$. For each n , determine the polynomial $p_n(x)$ that minimizes the Training Error, defined as the sum of the squares of the fitting errors at the 10 training data points. Plot the Training Error for the best fit polynomials of degree $n = 0, 1, 2, \dots, 10$.
 - Plot the Test Error of each of the polynomials obtained in (i) for a Test Set of $t = 100$ samples.
 - What value of n has the minimal Test Error?
 - Repeat (i,ii,iii) when $m = 20$.
- In this problem we will study the ℓ_2 -regularization method for the same datasets. Consider the regularized error with a regularization coefficient $\lambda > 0$:

$$\frac{1}{m} \left[\sum_{i=1}^m (y_i - p_n(x_i))^2 \right] + \lambda (w_0^2 + w_1^2 + \dots + w_n^2).$$

This is called Ridge Regression.

- For a range of values of $\lambda > 0$ ranging from very small to large, determine the polynomials $p_{10}^\lambda(x)$ that minimize the regularized error for each λ . Plot the Training Error of each of these polynomials as a function of λ . For the x -axis, you may want to have $\log \lambda$ rather than λ .
 - Plot the Test Error of each of the polynomials obtained in (i) for a Test Set of $t = 100$ samples.
 - What value of λ has the minimal Test Error? What is the corresponding polynomial? (In (i), you should choose a range of values of λ such that you can see the minimizing λ (i.e., the Test Error initially roughly decreases as λ increases and then roughly increases after a certain point).)
 - Repeat (i,ii,iii) when $m = 20$.
- ADDITIONAL PROBLEM - NOT GRADED.**
In this problem we will study the ℓ_1 -regularization method for the same datasets. Consider the regularized error with a regularization coefficient $\lambda > 0$:

$$\frac{1}{m} \left[\sum_{i=1}^m (y_i - p_n(x_i))^2 \right] + \lambda (|w_0| + |w_1| + \dots + |w_n|).$$

This is called LASSO (Least Absolute Shrinkage and Selection Operator).

(i) For a range of values of $\lambda > 0$ ranging from very small to large, determine the polynomials $p_{10}^\lambda(x)$ that minimize the regularized error for each λ . The minimization can be done by an optimization package such as `cvx`

https://www.cvxpy.org/examples/machine_learning/lasso_regression.html

Plot the Training Error of each of these polynomials as a function of λ . For the x -axis, you may want to have $\log \lambda$ rather than λ .

(ii) Plot the Test Error of each of the polynomials obtained in (i) for a Test Set of $t = 100$ samples.

(iii) What value of λ has the minimal Test Error? What is the corresponding polynomial? (In (i), you should choose a range of values of λ such that you can see the minimizing λ (i.e., the Test Error initially roughly decreases as λ increases and then roughly increases after a certain point).)

(iv) Repeat (i,ii,iii) when $m = 20$.