SANS2021-22-HW3-BiasVarianceTradeoff

jorge.garcia.vidal

December 2021

1 Introduction

In this HW you have to use the attached python program to investigate the Bias-Variance tradeoff that appear in regression problems. We will study the *ridge linear regression* estimators, that add a term in the optimization formula for finding the ML solution. The term penalizes large values for the components of θ. The function to minimize is:

$$\sum_{i=1}^{m}(t_i - \theta^t\phi(x_i))^2 + \lambda||\theta||_2^2 = \sum_{i=1}^{m}(t_i - \theta^t\phi(x_i))^2 + \lambda\theta^t\theta$$ understand the formula

where λ is a parameter that controls the penalization for large components of θ. Large values of the parameter produce biased estimators but with low variance. $\lambda = 0$ gives the ML solution. review ML

The regressors will be of the form: $\phi(x_i) = [1, x_i, x_i^2, .., x_i^{d-1}]^t$ for some value $d \in \{1, 2, ..\}$. What are regressors.

2 Case 1: Ordinary linear regression for a system with a linear response linear response?

Assume that the function we want to estimate is $f_\theta(x) = x$, i.e. it is a linear function of slope 1 and intercept coefficient 0, in the interval $[-4\pi/5, 4\pi/5]$. The variance of the noise that we add to the true value is $\sigma_\epsilon^2 = 2$. The size of the training set is $m = 5$.

Pendiente

intercept with the x axis inside the interval.

Run the program for $d = 2$ and $\lambda = 0$ (i.e. ordinary linear regression). Plot the histogram of the coefficient θ_1. Discuss the obtained results regarding the properties of Maximum Likelihood estimators. Experiment with different values of m and σ_ϵ^2 and discuss the results. To do

3 Case 2: Ridge regression for a system with a linear response

what is ridge regression?

Use the same assumptions as in the previous sections: $f_\theta(x) = x$, $\sigma_\epsilon^2 = 2$, and $m = 5$.
Run the program for $d = 2$ and $\lambda in\{0, 1, 10, 100\}$. Discuss the obtained results (Bias-variance tradeoff). Experiment with different values of m and σ_ϵ^2 and discuss the results.

To do

4 Case 3: Model selection and ordinary multiple linear regression

Assume that the function we want to estimate is $f_\theta(x) = sin(x)$ in the interval $[-4\pi/5, 4\pi/5]$. The variance of the noise that we add to the true value is $\sigma_\epsilon^2 = 2$. The size of the training set is $m = 10$.

Run the program for different polynomials degrees (set $d = 5$ and the program will run for polynomials from degree 1 to 4) and $\lambda = 0$ (i.e. ordinary linear regression). Take a look at the examples of regression functions you obtain for polynomial interpolations of different degrees. Plot the histograms of the coefficient θ_1. Plot the graph of MSE, Bias2 and Variance for different values of d. Discuss the obtained results regarding the properties of Maximum Likelihood estimators. Experiment with different values of m and σ_ϵ^2 and discuss the results.

To do

5 Case 4: Model selection and multiple Ridge regression

Assume that the function we want to estimate is $f_\theta(x) = sin(x)$ in the interval $[-4\pi/5, 4\pi/5]$. The variance of the noise that we add to the true value is $\sigma_\epsilon^2 = 1$. The size of the training set is $m = 50$.

Run the program for for different polynomials degrees (set $d = 5$ and the program will run for polynomials from degree 1 to 4) and $\lambda \in \{0, 1, 10, 100\}$. Discuss the obtained results regarding the properties of Maximum Likelihood estimators. Experiment with different values of m and σ_ϵ^2 and discuss the results.

To do