Homework 1 – Solutions

August 29, 2016

Due: September 5, 2017, 11:59 PM

Instructions

Your homework submission must cite any references used (including articles, books, code, websites, and personal communications). All solutions must be written in your own words, and you must program the algorithms yourself. If you do work with others, you must list the people you worked with. Submit your solutions as a PDF to the E-Learning at UF (http://elearning.ufl.edu/).

Your programs must be written in either MATLAB or Python. The relevant code to the problem should be in the PDF you turn in. If a problem involves programming, then the code should be shown as part of the solution to that problem. If you solve any problems by hand just digitize that page and submit it (make sure the problem is labeled).

If you have any questions address them to:

- Catia Silva (TA) – catiaspsilva@ufl.edu
- Sheng Zou (TA) – shengzou@ufl.edu

Question 1 – 10 points

Consider the polynomial curve fitting example discussed in class. As discussed, when the model order is *too* small, the training data is generally *underfit* and when the model order is *too* high, the result can *overfit* the training data. Write a small script of code that mimics our polynomial curve fitting function. The code should generate simulated data from the true function with added zero-mean Gaussian noise (with the true function assumed to be sine curve). The code should also generate a separate validation test data set generated in the same way. Then, after fitting the polynomial to the training data across a range of model orders and evaluated on both the training and testing data, your code should generate a plot similar to the one shown in Figure 1. Also, provide a discussion based on your plot about which model order, M, should be used to avoid over-training.

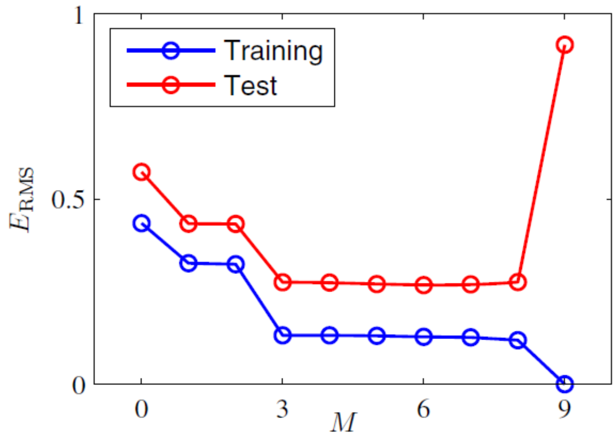

Figure 1: Figure 1: Figure 1.5 from the Bishop textbook. The x-axis corresponds to the root-mean-square error between the predicted and the treu value (on either the training data or test data sets). The x-axis corresponds to the model order.

A solution for this problem can be found in the jupyter notebook called *MyPolynomialRegression.ipynb*.

For the plot above, the model order that would avoid overfitting ranges from $M = 3$ to $M = 7$. But we always want the minimal model order possible in order to decrease computational costs. So the optimal order, in the RMS sense, to avoid overfitting and computational costs is $M = 3$.

Question 2 – 10 points

For each of the following problems, state whether or not the operation is defined (i.e., valid and can be computed) and, if it is defined, what is the size of the resulting answer. For all of the following problems, let $\mathbf{X}$ be an $M \times N$ matrix, $\mathbf{Y}$ be a $N \times N$ matrix, $\mathbf{a}$ be a $M \times 1$ vector, $\mathbf{b}$ be a $N \times 1$ vector and s be a scalar.

1. $\mathbf{XY}$ Defined. The result is a matrix of size $M \times N$.

2. $\mathbf{YX}$ Not defined.
3. $\mathbf{YX}^T$ Defined. The result is a matrix of size $N \times M$.
4. $\mathbf{aX}$ Not defined.
5. $\mathbf{a}^T\mathbf{X}$ Defined. The result is a vector of size $1 \times N$.
6. $\mathbf{aX}^T$ Not defined.
7. $\mathbf{a}^T\mathbf{b}$ Not defined.
8. $\mathbf{b}^T\mathbf{b}$ Defined. The result is a scalar.
9. $\mathbf{bb}^T$ Defined. The result is a matrix of size $N \times N$. (This is also called the *outer-product*).
10. $s\mathbf{X} + \mathbf{Y}$ Not defined.

Question 3 – 10 points

If $\mathbf{X}$ is a rank r matrix, show that the two square matrices $\mathbf{XX}^H$ and $\mathbf{X}^H\mathbf{X}$ have the same nonzero eigenvalues. (*Note: This is Problem 6.5 from the textbook*).
If $\mathbf{X}$ is a rank r matrix and suppose $\mathbf{X}$ is an $N \times M$ matrix, then its singular value decomposition can be written as:

$$\begin{aligned} \mathbf{X} &= \mathbf{U}_{N\times r}\Sigma_{k\times k}\mathbf{V}^H_{r\times M} \\ \mathbf{X}^H &= \mathbf{V}_{r\times M}\Sigma_{k\times k}\mathbf{U}^H_{N\times r} \end{aligned} \tag{1}$$

where $\mathbf{U}$ is an $N \times r$ real or complex unitary matrix, Σ is a $r \times r$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and $\mathbf{V}$ is an $r \times M$ real or complex unitary matrix. The diagonal entries of Σ are the singular values of $\mathbf{X}$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called the left-singular vectors and right-singular vectors of $\mathbf{X}$, respectively. Therefore,

$$\begin{aligned} \mathbf{XX}^H &= \mathbf{U}_{N\times r}\Sigma_{k\times k}\mathbf{V}^H_{r\times M}\mathbf{V}_{r\times M}\Sigma_{k\times k}\mathbf{U}^H_{N\times r} \\ &= \mathbf{U}_{N\times r}\Sigma_{k\times k}\left(\mathbf{V}^H_{r\times M}\mathbf{V}_{r\times M}\right)\Sigma_{k\times k}\mathbf{U}^H_{N\times r} \\ &= \mathbf{U}_{N\times r}\Sigma_{k\times k}\left(\mathbf{I}_{r\times r}\right)\Sigma_{k\times k}\mathbf{U}^H_{N\times r} \\ &= \mathbf{U}_{N\times r}\Sigma^2_{k\times k}\mathbf{U}^H_{N\times r} \end{aligned} \tag{2}$$

We can do the same for $\mathbf{X}^H\mathbf{X}$:

$$\begin{aligned} \mathbf{X}^H\mathbf{X} &= \mathbf{V}_{r\times M}\Sigma_{k\times k}\mathbf{U}^H_{N\times r}\mathbf{U}_{N\times r}\Sigma_{k\times k}\mathbf{V}^H_{r\times M} \\ &= \mathbf{V}_{r\times M}\Sigma_{k\times k}\left(\mathbf{U}^H_{N\times r}\mathbf{U}_{N\times r}\right)\Sigma_{k\times k}\mathbf{V}^H_{r\times M} \\ &= \mathbf{V}_{r\times M}\Sigma_{k\times k}\left(\mathbf{I}_{k\times k}\right)\Sigma_{k\times k}\mathbf{V}^H_{r\times M} \\ &= \mathbf{V}_{r\times M}\Sigma^2_{k\times k}\mathbf{V}^H_{r\times M} \end{aligned} \tag{3}$$

Thus, the non-zero elements of Σ (elements on its diagonal) or *eigenvalues* are the square roots of the nonzero singular values of $\mathbf{XX}^H$ and $\mathbf{X}^H\mathbf{X}$ and these singular values are the same.

Question 4 – 5 points

Consider $f(\mathbf{x}) = 3\mathbf{x}^T\mathbf{x} + 4\mathbf{y}^T\mathbf{x} - 1$ where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

1. What is $\frac{\partial f}{\partial \mathbf{x}}$? Show your work.

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial \left(3\mathbf{x}^T\mathbf{x} + 4\mathbf{y}^T\mathbf{x} - 1\right)}{\partial \mathbf{x}} \\ &= 3(\mathbf{x}^T + \mathbf{x}^T) + 4\mathbf{y}^T \\ &= 6\mathbf{x}^T + 4\mathbf{y}^T\end{aligned} \tag{4}$$

2. What is $\frac{\partial^2 f}{\partial \mathbf{x}^2}$? Show your work.

$$\begin{aligned}\frac{\partial f^2(\mathbf{x})}{\partial \mathbf{x}^2} &= \frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right) \\ &= \frac{\partial \left(6\mathbf{x}^T + 4\mathbf{y}^T\right)}{\partial \mathbf{x}} \\ &= \frac{\partial \left(6\mathbf{x}^T\mathbf{I} + 4\mathbf{y}^T\right)}{\partial \mathbf{x}}, \text{ where } I \text{ is the identity matrix of size } d \times d \\ &= 6\mathbf{I}\end{aligned} \tag{5}$$

Question 5 – 5 points

Consider $f(\mathbf{x}) = -10\mathbf{x}^T\mathbf{Q}\mathbf{x} + 4\mathbf{y}^T\mathbf{x} + 2$ where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\mathbf{Q} \in \mathbb{R}^{d\times d}$ and $\mathbf{Q}$ is symmetric.

1. What is $\frac{\partial f}{\partial \mathbf{x}}$? Show your work.

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial \left(-10\mathbf{x}^T\mathbf{Q}\mathbf{x} + 4\mathbf{y}^T\mathbf{x} + 2\right)}{\partial \mathbf{x}} \\ &= (-10\mathbf{Q}\mathbf{x})^T - 10\mathbf{x}^T\mathbf{Q} + 4\mathbf{y}^T \\ &= -10\mathbf{x}^T\mathbf{Q}^T - 10\mathbf{x}^T\mathbf{Q} + 4\mathbf{y}^T \\ &= -10\mathbf{x}^T\left(\mathbf{Q}^T + \mathbf{Q}\right) + 4\mathbf{y}^T, \ \mathbf{Q} \text{ is symmetric, i.e., } \mathbf{Q}^T = \mathbf{Q} \\ &= -10\mathbf{x}^T(2\mathbf{Q}) + 4\mathbf{y}^T \\ &= -20\mathbf{x}^T\mathbf{Q} + 4\mathbf{y}^T\end{aligned} \tag{6}$$

2. What is $\frac{\partial^2 f}{\partial \mathbf{x}^2}$? Show your work.

$$\begin{aligned}\frac{\partial f^2(\mathbf{x})}{\partial \mathbf{x}^2} &= \frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right) \\ &= \frac{\partial \left(-20\mathbf{x}^T\mathbf{Q} + 4\mathbf{y}^T\right)}{\partial \mathbf{x}} \\ &= -20\mathbf{Q}^T, \ \mathbf{Q} \text{ is symmetric, i.e., } \mathbf{Q}^T = \mathbf{Q} \\ &= -20\mathbf{Q}\end{aligned} \tag{7}$$